



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

이학박사학위논문

Percolation transitions in growing networks and simplicial complexes

성장하는 네트워크 및 단체 복합체에서의 여과
상전이

2021년 2월

서울대학교 대학원
물리·천문학부
오수민

Percolation transitions in growing networks and simplicial complexes

성장하는 네트워크 및 단체 복합체에서의 여과
상전이

지도교수 강병남

이 논문을 이학박사 학위논문으로 제출함

2021년 2월



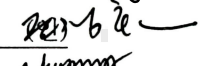


서울대학교 대학원

물리·천문학부

오수민

오수민의 이학박사 학위논문을 인준함

2021년 2월

위원장	최무영	
부위원장	강병남	
위원	조정호	
위원	백용주	
위원	손승우	

Abstract

Percolation transitions in growing networks and simplicial complexes

Soo Min Oh

Department of Physics and Astronomy

The Graduate School

Seoul National University

Percolation theory can explain the formation of a giant cluster in a system. About a decade ago, various types of local suppression rules were proposed to alter the transition types into explosive percolation (EP) transitions and these rules suppress the growth of larger clusters but support those of smaller ones. When the growth of the large cluster is locally suppressed in static networks whose total number of nodes in the system is fixed, explosive percolating behavior of the order parameter, which is the giant cluster size, can be observed but it remains continuous not discontinuous. However, the types of percolation transitions in static networks become hybrid when the growth of the large cluster is globally suppressed. In the real world, the system grows and we thus consider the growing network where the total number of nodes in the system increases with time. To investigate the properties of EP transitions in these types of systems, we first derive the rate equation

of the size distribution for various models. Our findings confirm that the order parameter exhibits the second-order phase transition under the local suppression rule, but the first-order phase transition under the global suppression rule. More realistically, considering that multi-body interactions in the real world can be explained by simplexes and degree distribution is scale free, we extend the growing networks to the growing scale-free simplicial complexes (GSFSC). In GSFSC under the suppression rules, the transition types are similar to those of the case of networks. Finally, we investigate the features of the percolation variables using supervised and unsupervised machine learning approach. Especially we confirm that features of parent node number and occupation number configurations are the giant cluster sizes and occupation probabilities, respectively, in the two-dimensional bond (site) percolation. Based on these results, we learned successfully classification and regression machines in networks by using supervised learning analysis. This work helps to understand the properties in general growing system and revealed that the machine learning approach is applicable to networks as well as lattices.

Keywords : Percolation transition, Explosive percolation transition, Discontinuous percolation transition, Rate equation, Generating function, Finite size scaling theory, Scale-free, Simplicial complex, Machine learning

Student Number : 2016-30097

Contents

Abstract	i
Contents	iii
List of Figures	vi
List of Tables	xxvi
1. Introduction	1
1.1 Percolation	1
1.2 Two models: the growing and the static network models . . .	2
1.2.1 Callaway network model	2
1.2.2 Erdős-Rényi (ER) network model	3
1.3 Suppression effects	4
1.4 Growing simplicial complexes	4
2. Local suppression effects in growing networks	6
2.1 Minimal rule	6
2.1.1 Model	6
2.1.2 Rate equation approach for the cluster size distribution	7
2.1.3 Monte Carlo simulations	20
2.1.4 Comparison of $n_s(p)$ for growing network models .	25
2.1.5 Analysis	26
2.2 da Costa rule	27

2.2.1	Model and rate equation	27
2.2.2	Scaling relations of critical exponents	29
2.2.3	Analytic solution of the transition point	33
2.2.4	Numerical solutions of the rate equation	37
2.2.5	Analysis	41
3.	Global suppression effects in growing networks	45
3.1	Model: r -GRN model	45
3.2	Cluster size distribution $n_s(p)$	47
3.3	Two transition points, p_b and p_c	51
3.3.1	For finite S_R	53
3.3.2	For infinite S_R	55
3.4	$\tau(p)$ in the critical region and total number of clusters	58
3.5	Universal behavior	60
3.6	Analysis	62
4.	Growing scale-free simplicial complexes	69
4.1	Models and rate equations	69
4.2	Degree and facet degree distributions	75
4.3	Poisson distribution	79
4.4	Giant cluster size	84
4.5	Size distribution	89
4.6	The case of $(d + 1)$ -sided polygon	91
4.7	Analysis	91
5.	Machine learning approaches	93

5.1	Motivation	93
5.2	Model and method	94
5.3	Results	96
5.3.1	Relation between parent node number and giant cluster size	97
5.3.2	ER network	98
5.3.3	r -ER network	101
5.4	Analysis	105
6.	Conclusion	107
	Appendices	110
	Appendix A. r-GRN model	111
	A.1 Rate equations of the r -GRN model	111
	Appendix B. Unsupervised learning of percolation transitions	
	in two-dimensional lattices	114
	B.1 Feature extraction	114
	Bibliography	116
	Abstract in Korean	123

List of figures

1.1.	(Color online) Illustrations of d -simplexes for $d =$ (a) 1, (b) 2, (c) 3, and (d) 4. In each figure, 0-simplex is represented by a open circle.	5
2.2.	(Color online) For growing networks with $m = 3$, plot of $n_s(p)$ vs s at $p = p_c$ (blue solid line), $p > p_c$ (red dashed curves) and $p < p_c$ (black solid curves) based on numerical values obtained from the rate equation. The transition point p_c is $p_c = 0.413842(1)$, and the exponent τ is approximately 2.5. The black dashed line is a guide line with slope -2.5	10
2.3.	(Color online) For growing networks with $m = 3$, plot of $n_s(p)s^\tau$ versus $s p - p_c ^{1/\sigma}$ for different values of p when (a) $p < p_c$ and (b) $p > p_c$. Data for different p values are well collapsed onto a single curve by choosing $\sigma = 0.720(2)$ and $\tau = 2.500(1)$	10
2.4.	(Color online) For growing networks with $m = 3$, plot of $G(p)$ vs p . The data points are obtained from the rate equation. Inset: The dashed line is a guide line with slope $0.694(2)$	11

- 2.5. (Color online) For growing model with $m = 3$, plot of the susceptibility, that is mean cluster size, as a function of p . The data points are obtained from the rate equation. Insets : Double logarithmic plots of $\langle s \rangle$ versus $|p - p_c|$ for $p < p_c$ (left) and $p > p_c$ (right). The dashed lines are guide lines with slope $-0.696(3)$. . . 12
- 2.6. (Color online) For growing model with general m , (a) empirical plot of $1 - p_c$ versus m . Data points of $1 - p_c$ for different values of m behave like the formula $1.81/m$. (b) Empirical plot of β vs m . Data points of β for different values of m behave like the formula $\beta = 1/(m - 1.56)$. Note that the error bars are smaller than the symbol sizes. 13
- 2.7. (Color online) For growing networks with general m , formula testing for the exponents of (a) the cluster size distribution τ , (b) the characteristic cluster size σ , and (c) the mean cluster size γ , where the numerical data are obtained from the rate equation. Data are fit reasonably to the straight line predicted by the formula, and the error bars are smaller than the symbol sizes. . . 14

- 2.8. (Color online) For static networks with $m = 3$, plot of $n_s(t)$ vs s at $t = t_c$ (blue solid line), $t > t_c$ (red dashed curves) and $t < t_c$ (black solid curves) based on the numerical values obtained from the rate equation. The transition point t_c is determined as $t_c = 0.849130(1)$ and the exponent τ is determined as $\tau = 2.105(5)$. Black dashed line is a guideline with slope -2.105 16
- 2.9. (Color online) For static networks with $m = 3$, scaling plot of $n_s(t)s^\tau$ versus $s|t - t_c|^{1/\sigma}$ for different t that are (a) less and (b) greater than t_c . Taking $\tau = 2.105(5)$ and $\sigma = 0.790(1)$, the data for different t values look collapsed onto a single curve. 16
- 2.10. (Color online) For static networks with $m = 3$, plot of the order parameter $G(t)$ as a function of t . Inset : The dashed line is a guide line with slope $0.133(1)$ 17
- 2.11. (Color online) For static networks with $m = 3$, plot of $\langle s \rangle$ as a function of t . Inset : Plot of the susceptibility, the mean cluster size as a function of t for $t > t_c$ (right) and $t < t_c$ (left). The dashed lines are guide lines with slope $-1.131(6)$ 17

- 2.12. (Color online) For static networks, (a) plot of $1 - t_c$ versus m on a semi-logarithmic scale. (b) Plot of the estimated values of the exponent β for general m versus m on a semi-logarithmic scale. Asymptotically, the data points likely lie on a straight line. The error bars are smaller than the symbol sizes. 19
- 2.13. (Color online) For static networks, formula testing for the exponents of (a) the cluster size distribution τ , (b) the characteristic cluster size σ , and (c) the mean cluster size γ , where the numerical data is obtained from the rate equation. Note that the data fit reasonably to the straight line predicted by the formula. 19
- 2.14. (Color online) To obtain τ and σ by Monte Carlo simulations for growing networks, scaling plot of $n_s(p)s^\tau$ versus $s|p - p_c|^{1/\sigma}$. Data are collapsed onto a single curve by choosing $\tau = 2.5$ and $\sigma = 0.72$ for (a) $p < p_c$ and (b) $p > p_c$ 20
- 2.15. (Color online) To obtain β by Monte Carlo simulations for growing networks, scaling plot of $GN^{\beta/\bar{\nu}}$ versus $(p - p_c)N^{1/\bar{\nu}}$ for system sizes $N/10^4 = 2^3 - 2^{10}$. Data are collapsed onto a single curve with the values of $1/\bar{\nu} = 0.35(3)$ and $\beta/\bar{\nu} = 0.24(3)$ 21

- 2.16. (Color online) To obtain the exponent γ using Monte Carlo simulations for growing networks, data collapse plot of $\langle s \rangle N^{-\gamma/\bar{\nu}}$ versus $(p - p_c)N^{1/\bar{\nu}}$ for the system sizes $N/10^4 = 2^6 - 2^{10}$ in growing networks. The exponent values are $\gamma = 0.696$ and $1/\bar{\nu} = 0.35$ 21
- 2.17. (Color online) To obtain the exponents τ and σ using Monte Carlo simulations for static networks, data collapse plots of the rescaled cluster size distribution $n_s(t)s^\tau$ versus $s|t - t_c|^{1/\sigma}$ for different time steps when (a) $t < t_c$ and (b) $t > t_c$, where $\tau = 2.105$ and $\sigma = 0.79$. 23
- 2.18. (Color online) To obtain β using Monte Carlo simulations for static networks, data collapse plot of $GN^{\beta/\bar{\nu}}$ versus $(t - t_c)N^{1/\bar{\nu}}$ for system sizes $N/10^4 = 2^0 - 2^{10}$. Data for different values of N are systematically collapsed near the transition point by taking $1/\bar{\nu} = 0.45$ and $\beta/\bar{\nu} = 0.06$ 24
- 2.19. (Color online) To obtain γ using Monte Carlo simulations for static networks, data collapse plot of $\langle s \rangle N^{-\gamma/\bar{\nu}}$ versus $(p - p_c)N^{1/\bar{\nu}}$ for system sizes $N/10^4 = 2^7 - 2^{10}$, where the exponent values $\gamma = 1.133$ and $1/\bar{\nu} = 0.45$ are used. 24

- 2.20. (Color online) For growing network with $m = 2$, plot of $n_s(p)$ versus s at $p = p_c$ (blue solid line), $p > p_c$ (red dashed line), and $p < p_c$ (black solid line) based on the numerical values obtained from the rate equation. The transition point is $p_c = 0.125$. For $p \leq p_c$, $n_s(p)$ decays in a power-law manner, indicating that the transition is infinite-order. 25
- 2.21. (Color online) Schematic illustration of our model for $m = 3$. Nodes are represented by solid circles. After a new node (red dotted open circle) is added to the system, two sets of m nodes (solid open circles) are randomly selected from distinct clusters. For each set represented by a shaded ellipse, the node belonging to the smallest cluster (represented by \oplus) is chosen. The two nodes \oplus are connected with probability p by a link (dashed line). 28

- 2.22. (Color online) Plot of $f(0)$ versus $1/s_0$ for $m = 2$ with s_0 for truncated cluster sizes $s^*/10^3 = 2^{11}, 2^{12}, 2^{13}, 2^{14}$, and 2^{15} (red dashed curves from bottom to top), and 2^{16} (red solid curve). Inset: The corresponding plot of $f(0) - f^*(0)$ versus s^* at $p = 0.51515$ around the transition point p_c , where $f^*(0)$ is the minimum value of $f(0)$ for a given s^* . Data points are for trial values of $f^*(0) = 0.213$ (●), 0.217 (■), and 0.223 (▲). The black dashed line is a guide line and $f(0)$ is estimated as $0.217(1)$ for $m = 2$ 36
- 2.23. (Color online) (a) Plot of $1 - p_c$ versus m . Data points obtained from the rate equation seem to be fitted by the formula $1 - p_c = 1.04/m$. (b) Plot of β versus m . Data points obtained from the rate equation are fitted by the formula $\beta = 1/(2m - 1.87)$. The critical exponent β decreases algebraically with increasing m 38
- 2.24. (Color online) Test of the formulas for the exponents (a) $\tau = 2 + 1/(2m - 1)$, (b) $1/\sigma = (2m - 1)\beta$, (c) $\gamma_P = 2(m - 1)\beta$, and (d) $\gamma_Q = (m - 1)\beta$ using numerical data obtained from the rate equation. The error bars are presented in (a), (b), and (d). Data are fitted to the straight solid line, $y = x$, following Eqs. (2.16)–(2.19) in our growing network model. 39

- 2.25. (Color online) Finite-size scaling of $GN^{\beta/\bar{\nu}}$ versus $(p - p_c)N^{1/\bar{\nu}}$ from Monte Carlo simulations with more than 10^4 realizations for system sizes $N/10^4 = 2^3, \dots, 2^{10}$ for $m = 2$. Data collapse onto a single curve with $p_c = 0.515(1)$, $1/\bar{\nu} = 0.383(10)$, and $\beta/\bar{\nu} = 0.175(10)$. One then get $\beta = 0.458(38)$ 40
- 2.26. (Color online) Schematic plots of the order parameter G and the inverse first moment of $P_m(s, p)$, $1/\langle s \rangle_p$, for the (a) GRN, (b) d -GRN/ m -GRN, (c) ER, and (d) d -ER/ m -ER models. Schematic plots for the m -GRN and m -ER models are very similar to those for the d -GRN and d -ER models. The only difference between the two suppression rules is that twice as many nodes are selected under the da Costa rule as under the minimal rule. Thus, $m = 2$ under the minimal rule and $m = 1$ under the da Costa rule are reduced to the GRN model in growing networks and the ER model in static networks, respectively. 42

3.27. Schematic illustration of the r -GRN model with $g = 0.4$. Nodes are represented by circles. $\mathbf{R}(t)$ is represented by light gray region and $\mathbf{R}^c(t)$ is done by dark grey region. Solid line between two nodes represents a link. In (a), the system starts at five clusters with sizes $(1, 1, 2, 2, 4)$, respectively, and the total number of nodes $N(t) = 10$ with $S_R = 2$ and $\lceil gN \rceil = 4$ at $t = 9$. After one time step, a new node (red open circle) is added, and $N(t)$ becomes 11 with $S_R = 2$ and $\lceil gN \rceil = 5$ at $t = 10$. Two isolated nodes (filled light grey) in \mathbf{R} and are merged. (b) Next, a new node is added, and so $N(t) = 12$, $S_R = 2$, and $\lceil gN \rceil = 5$ at $t = 11$. In this case, one node of the cluster of size two in set \mathbf{R} moves to set \mathbf{R}^c . The newly added node is merged with the cluster of size two in \mathbf{R} . This cluster moves to set \mathbf{R}^c and the cluster of size two on the boundary moves to \mathbf{R} . (c) Next, a new node is added with $N(t) = 13$, $S_R = 2$, and $\lceil gN \rceil = 6$ at $t = 12$. Selected nodes are not connected with probability $1 - p$. (d) Next, a new node is added with $N(t) = 14$, $S_R = 2$, and $\lceil gN \rceil = 6$ at $t = 13$. And just one node of the cluster of size two in set \mathbf{R} moves to set \mathbf{R}^c . A cluster of size two in \mathbf{R} and the cluster of size three in \mathbf{R}^c are merged, belonging to \mathbf{R}^c . The cluster of size four in \mathbf{R}^c lies on the boundary and the cluster of size two on the boundary moves to \mathbf{R} 46

3.28. Cluster size distribution $n_s(p)$ as a function of s and p for given g : In this case, $g = 0.4$ are taken. (a) 3D plot of $n_s(p)$ as a function of s and p . Plots (b)-(d) are obtained with several fixed p for $n_s(p)$. (b) For $p < p_b$, $n_s(p)$ follows $\sim s^{-\tau}$ with $\tau > 3$. The slope of the dotted guide line is -3 . Solid lines are obtained for $p = 0.472576 \approx p_b, 0.4, 0.3, 0.2$, and 0.1 from right to left. (c) For $p_b \leq p < p_c$, in the small-cluster-size region, $n_s(p)$ decays exponentially up to S_R and then exhibits power-law decay behavior with $2 < \tau \leq 3$. Solid, dashed, and dashed-dotted lines represent for p_{S_R} with $S_R = 2, 10$ and 25 , respectively. Dotted line is a guide line with slope of -2 . (d) For $p \geq p_c$, $n_s(p)$ exponentially decay. Solid curves represent $n_s(p)$ for $p = 0.6596, 0.7, 0.8, 0.9$, and 1.0 from right to left. Dotted curve is an exponentially decaying guide curve. Plots (e)-(h) are obtained with several values of s for $n_s(p)$. (e) Plot of $n_1(p)$ versus p . A crossover exists at p_1 . (f) Plot of $n_2(p)$ versus p . Two crossover behaviors occur at p_1 and p_2 , where $p_1 < p_2$. (g) and (h) Plots of $n_3(p)$ and $n_4(p)$ versus p , respectively. Symbols represent simulation results, and solid lines are analytical results. Dotted vertical lines represent p_{S_R} for $S_R = 1, 2, 3$, and 4 at $p_{S_R=1} = 0.428571$, $p_{S_R=2} = 0.565302$, $p_{S_R=3} = 0.612016$, and $p_{S_R=4} = 0.632728$ 49

- 3.29. Plot of $p_\infty - p_{S_R}$ versus S_R for $g = 0.4$. When $p_\infty = 0.65948(1)$, a power-law decay appears. 51
- 3.30. Semi-log plot of $p^* - p_{S_R}$ versus S_R . $p^* - p_{S_R}$ decays exponentially as S_R increases but it's always larger than zero for finite S_R values. So p^* is larger than p_{S_R} for any finite S_R 54
- 3.31. Plot of G and $1/\langle s \rangle$ as a function of p . For $g = 0.2$ in (a) and (d), $g = 0.4$ in (b) and (e), and $g = 0.6$ in (c) and (f), respectively. Symbols represent the simulation results for $N = 10^4$ (\circ), 10^5 (\triangle), 10^6 (\square), and 10^7 (\diamond). Each data point was averaged over 10^3 times. The solid (red) lines are calculated from $f(1)$ and $f'(1)$ for G and $\langle s \rangle$, respectively. The two vertical dotted lines represent p_b and p_c ($p_b < p_c$). 57
- 3.32. (a) and (b) show the phase diagrams of the r -GRN model and r -PIN model with $\delta = 0.7$, respectively. Symbols \triangle and \circ represent p_b and p_c . $n_s(p)$ decays following a power law with $\tau > 3$ in the infinite-order-type critical region and $2 < \tau < 3$ in the second-order-type critical region. Thus, the mean cluster size is finite and diverges, respectively. As g approaches one, two phase boundaries converge to the conventional transition point $p_c = 1/8$ of the GRN model, represented by \blacksquare , and $p_c = 0.12(1)$ in the PIN model with $\delta = 0.7$, represented by \blacklozenge 57

- 3.33. Plot of τ versus p for different g . τ becomes two as p approaches p_c for any g . The black dashed curve is a guide curve representing $1 + 1/p$, which is obtained from the limiting case $S_R = 1$, i.e., $g \rightarrow 0$ 59
- 3.34. Plot of the total number of clusters n_{cl} versus p for $g = 0.4$. The red solid line is obtained from the rate equation integrating the cluster size distribution. The open circles represent the numerical simulation data for $N = 10^6$, averaged over 10^4 configurations. The black vertical dotted line represents p_c for $g = 0.4$. . . 59

3.35. Plots of the cluster size distribution $n_s(p)$ of the r -PIN model as a function of s in different p regions. We binned simulation data logarithmically for $N = 2^{14} \times 10^4$ averaged over 10^3 configurations. $g = 0.4$ and $\delta = 0.7$ are taken. Three cases of $n_s(p)$ are distinguished: (a) For $p < p_b$, $n_s(p)$ asymptotically follows the power law $\sim s^{-\tau}$ with $\tau > 3$. The slope of the dotted guide line is -3 . Solid lines are obtained for $p = 0.29 \approx p_b, 0.25, 0.20, 0.15, 0.10$, and 0.05 from right to left. (b) For $p_b \leq p < p_c$, in the small-cluster-size region, $n_s(p)$ decays exponentially and then exhibits power-law decay behavior with $2 < \tau \leq 3$. Solid (black), dashed (red), and dashed-dotted (blue) lines represent p_{S_R} , where $S_R = 2$ ($p = 0.29$), 3 ($p = 0.35$), and 17 ($p = 0.423$), respectively. Two dotted lines are guidelines with slopes of -2 and -3 . (c) For $p \geq p_c$, $n_s(p)$ for finite clusters shows exponentially decaying distributions. Solid curves represent $p = 0.43, 0.50, 0.60, 0.75$, and 0.90 from right to left. Dotted curve is an exponentially decaying guide curve. 61

3.36.	Plot of G and $1/\langle s \rangle$ for the r -PIN model as a function of p for $g = 0.4$ in (a) and (c), and for $g = 0.6$ in (b) and (d), respectively. Symbols represent numerical simulation data for $N = 2^2 \times 10^4$ (\bigcirc), $2^8 \times 10^4$ (\triangle), and $2^{14} \times 10^4$ (\square). Each data point was averaged over 10^2 configurations. The two vertical dotted lines represent p_b and $p_c > p_b$	62
3.37.	Schematic plots of the order parameter G and the inverse mean cluster size $1/\langle s \rangle$ for the (a) ER, (b) r -ER, (c) GRN, and (d) r -GRN models.	65
4.38.	(Color online) Phase diagram of growing scale-free simplicial complexes for (a) $d = 1$ (link), (b) 2 (triangle), (c) 3 (tetrahedron), and (d) 4 (5-cell) in (p, a) plane, where p and a are the probability, which represents the facet density, and the initial attractiveness, respectively. Each solid line represents the critical line satisfying $p = p_c(d, a)$. The giant cluster exists in the region above each critical line for $p > p_c(d, a)$, while it is absent in the region below each critical line for $p < p_c(d, a)$	75

- 4.39. (Color online) Plots of τ versus p in growing scale-free simplicial complexes for (a) $d = 1$ (link), (b) 2 (triangle), (c) 3 (tetrahedron), and (d) 4 (5-cell). Red ($a = 0.5$), green ($a = 1.0$), and blue ($a = 2.0$) symbols represent the results obtained from measures by solving numerically the rate equation of size distribution. Each solid line is the analytic solution $\tau = 3 + 4\sqrt{-\alpha}/(1 - 2\sqrt{-\alpha})$. For each plot, dotted lines from the left represent the critical points for $a = 0.5$, 1.0, and 2.0. 91

5.40.	(Color online) Schematic illustration of the machine learning (ML) process in our convolutional neural networks. We first introduce the ER and r -ER networks starting at $N = 20$ isolated nodes and the parents of all nodes are themselves at the beginning. When two clusters are merged as the network grows, parents of all nodes in cluster with smaller number of parents node are changed into parents of all node in other cluster before the merger. For given time t , the parents node numbers of all nodes are embedded into sites on the one dimensional lattice of length size 20. The feature of this embedded information is learned through successive convolutional and pooling layers twice with several filters. kernel size of convolutional layer is 4 and pool size of pooling layer is 4. The values of sites in the last pooling layer are flattened and goes through a fully connected network to classify or regress the information in the network. For the classification machine, the activation functions of output layer and the other layer in FCN are softmax and the scaled exponential linear units (SELUs) functions, respectively. For the regression machines, the activation function of all layer in FCN are SELU functions. Loss function is mean squared error and optimizer is Adam for training our machines.	95
-------	--	----

- 5.41. (Color online) For data from Monte Carlo (MC) simulations, plots of the averaged parent node numbers of nodes in the giant cluster versus the giant cluster size for the system size $N/10^2 = 2^7$ with 1000 realizations for (a) ER and (b) r -ER networks. The color of data point represents the link density. The closer link density t to 0 the more blue it becomes, meanwhile the closer its value to 1.2 the more red it becomes. When $t = t_c$, the color is light gray. Spearman correlation coefficients, between the parent node numbers of nodes in the giant cluster and the giant cluster sizes, are 0.89 for ER and 0.90 for r -ER networks. Hence, our results confirm that the probability, that parent node number of the node in the giant cluster is the largest, increases with time t 98
- 5.42. (Color online) For the classification machines of ER networks, plots of (a) output values (y_1, y_2) versus the link density t . Training regions are shaded in orange. (b) Plot of $t_c - t_c'(N)$ versus N , where the asymptotic behavior of $t_c'(N)$ follow $t_c - t_c'(N) \sim N^{-1/\bar{v}'}$ with $t_c = 0.50(1)$ $\bar{v}' = 3.0(1)$. (c) Plot of (y_1, y_2) versus $(t_c - t)N^{1/\bar{v}'}$ and all data are collapsed to a single curve with the estimated t_c and \bar{v}' 99

- 5.43. (Color online) Plots of the retrieved values G_{ML} from regression machines versus the giant cluster sizes G_{MC} from Monte Carlo (MC) simulations for (a) training and (b) test results in ER networks for the system size $N/10^2 = 2^7$. In (a), the pixel of (0,0) has been enlarged to make it more visible. (c) The corresponding plot of the giant cluster size (black solid line) from MC simulation versus t , compared to the retrieved value (red open circles) from the regression machine versus t 100
- 5.44. (Color online) (a) Plot of $t_c(N) - t_c$ versus N for $N/10^2 = 2^2 - 2^7$. The value of $t_c(N)$ is defined as the link density where $dG(t)/dt$ is the maximum. Assuming that $t_c(N)$ follow $t_c(N) - t_c \sim N^{-1/\bar{\nu}^*}$, one can get $t_c = 0.50(1)$ and $\bar{\nu}^* = 3.0(1)$. From the finite size scaling formula $G(t) = N^{-\beta/\bar{\nu}} f((t - t_c)N^{1/\bar{\nu}})$, data points for all system sizes are collapsed to a single curve with $t_c = 0.5(1)$ and $\bar{\nu} = 3.0(1)$ in (b) plot of $GN^{\beta/\bar{\nu}}$ versus $(t - t_c)N^{1/\bar{\nu}}$. All data points above are obtained from our trained regression machines for ER network. . . . 100

- 5.45. (Color online) For the classification machines of r -ER networks, plots of (a) output values (y_1, y_2) versus the link density t . Training regions are shaded in orange. (b) Plot of $t_c - t'_c(N)$ versus N , where the asymptotic behavior of $t'_c(N)$ follow $t_c - t'_c(N) \sim N^{-1/\bar{v}'}$ with $t_c = 0.725(2)$ $\bar{v}' = 1.8(1)$. (c) plot of (y_1, y_2) versus $(t_c - t)N^{1/\bar{v}'}$ and all data are collapsed to a single curve with the estimated t_c and \bar{v}' 102
- 5.46. (Color online) Plots of the retrieved value G_{ML} from regression machines versus the giant cluster sizes G_{MC} from MC simulation for (a) training and (b) test results in r -ER networks for the system size $N/10^2 = 2^7$. In (a), the pixel of (0,0) has been enlarged to make it more visible. (c) The corresponding plot of the giant cluster size (black solid line) from MC simulation versus t , compared to the retrieved value (red open circles) from the regression machine versus t 103

5.47. (Color online) (a) Plot of $G_0 - G_0(N)$ versus N for $g = 0.8$ for $N/10^2 = 2^5 - 2^7$, including additional sizes of $N/10^2 = 23, 45$, and 90 . When $G_0 = 0.29(3)$, a power-law decay appears. (b) Plot of $t_c - t_c(N)$ versus N for $N/10^2 = 2^2 - 2^7$. The value of $t_c(N)$ is defined as t intercept of the tangent of $G(t)$ at the point where $dG(t)/dt$ is the maximum. Assuming that $t_c(N)$ follow $t_c - t_c(N) \sim N^{-1/\bar{\nu}^*}$, one can get $t_c = 0.725(2)$ and $\bar{\nu}^* = 1.8(1)$. From the finite size scaling formula $G(t) - G_0 = N^{-\beta/\bar{\nu}} f((t - t_c)N^{1/\bar{\nu}})$, data points for all system sizes are collapsed to a single curve with $t_c = 0.725(2)$, $G_0 = 0.29(3)$, $\beta = 0.32(3)$ and $\bar{\nu} = 1.8(1)$ in (c) plot of $(G - G_0)N^{\beta/\bar{\nu}}$ versus $(t - t_c)N^{1/\bar{\nu}}$. All data above are obtained from our trained regression machines for r -ER. 103

List of tables

- 2.1. Numerical estimates of the percolation threshold p_c , exponent of the cluster size distribution τ , exponent of the characteristic cluster size σ , exponent of the order parameter β , and exponent of the susceptibility γ of the growing network model for $m = 3, \dots, 10$. τ^* and β^* were obtained from $\tau^* = 2 + 1/(m-1)$ and $\beta^* \approx 1/(m-1.56)$, respectively. 14
- 2.2. Numerical estimates of the percolation threshold t_c , exponent of the cluster size distribution τ , exponent of the characteristic cluster size σ , exponent of the order parameter β , and exponent of the susceptibility γ of the static network model for $m = 2, \dots, 5$. Subsequently, τ^* and β^* were obtained from $\tau^* = 2 + \beta/[1 + (m-1)\beta]$ and $\beta^* \approx 0.465 \exp(-0.70m)$, respectively. 18

2.3.	Numerical estimates of the percolation threshold p_c , the exponent of the order parameter β , the exponent of the cluster size distribution τ , the exponent of the characteristic cluster size σ , and the exponents of the susceptibility γ_P , γ_Q of growing network models for $m = 2$ to 10. The transition point p_c obtained from the rate equations is compared with p_c^* , which was analytically solved using the scaling functions. $f(0)$ is the coefficient of the leading term of $P_m(s, p) \approx f(0)s^{-\tau}$ for large s . We confirm that p_c and p_c^* are consistent with each other within errors.	39
2.4.	Values of the upper critical dimension d_u obtained from the hyperscaling relation in growing networks for $m = 2 - 5$. The correlation volume exponent $d_u \nu$ for the mean-field theory values $\nu = 1/2$ and numerically estimated $\bar{\nu}$ from the finite-size scaling approach is consistent with simulation data within errors. The network is grown to $N = 2^{10} \times 10^4$, and the ensemble average is taken over more than 10^4 samples for each m	40

2.5.	Transition point $p_c(t_c)$; the order-parameter exponent β ; the scaling relations of the critical exponents τ , $1/\sigma$, and γ_p ; and the upper critical dimension d_u in growing (static) networks under two different suppression rules for candidate node selection. The minimal rule for the growing and static networks and the da Costa rule in the static networks are compared. The difference between these two suppression rules is that twice as many nodes are selected under the da Costa rule as under the minimal rule.	43
3.6.	Values of p_{S_R} as a function of S_R for $g = 0.4$	50
3.7.	Numerical estimates of the transition points p_b and p_c . The critical exponents τ are calculated at $p = p_b$ and p_c for $g = 0.1 - 0.9$. We note that the exponent τ at p_c becomes difficult to obtain as g approaches one.	52
3.8.	Comparison of the BKT transitions between in thermal systems and in percolations of the growing networks. . .	68

- 5.9. Comparison of the values of the critical point t_c , critical exponent β , jump size G_0 , correlation exponents $\bar{\nu}$ and $\bar{\nu}^*$ obtained from MC simulation and ML approach in Erdős-Rényi (ER) and restricted-ER (r -ER) networks. All values are consistent with each other. The critical exponents $\bar{\nu}$ is defined in the finite size scaling formula for the giant cluster size G in results of MC simulations and regression machines in ER and r -ER. The other exponent $\bar{\nu}^*$ is defined in the formula $|t_c - t_c(N)| \sim N^{-\bar{\nu}^*}$. We remark that $t_c(N)$ is defined as the point where dG/dt is the maximum in ER whereas it is defined as t intercept of the tangent of $G(t)$ at the point where $dG(t)/dt$ is the maximum in r -ER. Additionally, we define $\bar{\nu}'$ in $t_c - t_c'(N) \sim N^{-\bar{\nu}'}$, where $t_c'(N)$ is defined as the intersection point of output values (y_1, y_2) in classification machines in both ER and r -ER networks. The results from ML approaches are consistent with those from MC analysis. 104
- 5.10. In test data for regression machines, the Pearson correlation coefficient R and root mean square error (RMSE), bias between the giant cluster sizes from regression machines and MC simulations for the system size $N/10^2 = 2^2 - 2^7$ in each ER and r -ER ($g = 0.8$) networks. \bar{G}_{ML} and \bar{G}_{MC} represent the averaged giant cluster size for regression and MC simulation results, respectively. . . . 104

B.11.	The features, extracted by unsupervised learning methods, of percolation variables including site (bond) occupation number, degree, adjacency matrix, Laplacian matrix and parent node number for two-dimensional site (bond) percolation.	115
-------	--	-----

Chapter 1

Introduction

1.1 Percolation

Percolation is a simple yet basic model for understanding the emergence of a giant component as links are occupied with a certain probability between each pair of nodes in a system [1, 2]. This simple model has been applied to a variety of real-world phenomena such as the sol-gel transition [3–6], spreading of epidemic diseases [7–10], and the metal-insulator transition [11]. Conventionally, a percolation transition is second-order [1, 2]; however, interest in other types of percolation transitions such as first-order [12], infinite-order [13, 14] or mixed-order [15] phase transitions has increased recently. This trend has been triggered by the explosive percolation (EP) model [16] and an cascading failure model in interdependent networks [17, 18].

An EP model was introduced aiming to generate a discontinuous percolation transition, in which two potential edges are chosen randomly, and then an actual connection is made by the edge that produces the smaller component. After extensive researches were performed, it turned out that the EP model undergoes a second-order transition in the thermodynamic limit [19, 20]. However, the EP transition can be discontinuous under sufficient global information [21]. The critical exponent β of the order parameter

for the original EP model is extremely small [19], implying that the order parameter increases drastically at a transition point. Moreover, transition properties of the EP models are unconventional [22]. Extensive researches were performed mainly in static networks, in which the number of nodes are always fixed from the beginning. However, real-world phenomena related to such a drastic increase of the order parameter can be observed in growing social networks [23]. Nevertheless, the EP transition in growing networks has not been investigated in detail [24, 25]. Thus, we aim to investigate an EP transition in growing networks, based on our previous researches in [26–28]. In this thesis, we investigate the percolation transitions in growing networks and simplicial complexes, and we finally expect that our studies help the development of researches in complex networks.

1.2 Two models: the growing and the static network models

There are many percolation models in lattice, networks, etcetera but in this paper we focused on the percolation in networks. Networks consist of lines(edges) and nodes(vertices). In this section we introduce Erdős-Rényi (ER) and Callaway networks model which are classical percolation models in networks.

1.2.1 Callaway network model

Callaway networks model, which is the growing random networks (GRN), is introduced in 2001 [14] where the giant cluster size exhibits the

infinite-order phase transitions. This model starts at the N_0 isolated nodes and a new node is added to the system every time step t . Sequentially, two nodes are randomly selected and they are linked with probability p .

In this system, the order parameter, i.e., the relative giant cluster size, $G(p)$ is zero for $p < p_c$ and increases continuously for $p > p_c$ in the essentially singular form

$$G(p) \sim \exp(-a/\sqrt{p-p_c}), \quad (1.1)$$

where a is a positive constant and the transition point p_c is $p_c = 1/8$. Thus, the PT is infinite-order. In this case, the cluster size distribution $n_s(p)$ follows a power law $n_s \sim s^{-\tau}$ without any exponential cutoff in the entire region of $p < p_c$ [13, 14, 26, 29]. Thus, the region $p < p_c$ is often referred to as the critical region. The exponent τ decreases with increasing p and approaches $\tau = 3$ as $p \rightarrow p_c$ from below [13, 29]. Thus, the mean cluster size, $\langle s \rangle \equiv \sum_s s^2 n_s$, is finite for $p \leq p_c$. Moreover, for $p > p_c$, $n_s(p)$ of finite clusters decays exponentially. Thus, $\langle s \rangle$ is finite. These properties of a PT of growing networks are different from those of a second-order PT of static networks [2, 14, 24–26, 30]. This model can be extended to the growing scale-free networks introduced by Dorogovtsev in Refs. [29, 31, 32], as real networks are scale-free networks [33–35]

1.2.2 Erdős-Rényi (ER) network model

ER networks model is introduced in 1959 [36] is one of the percolation models which undergoes continuous phase transition. This ER model is the

growth network model with randomly selecting two nodes to be linked at each time t and the total number of nodes N is fixed. And there is no giant cluster if the average degree $\langle k \rangle$ is less than 1 in the limit $N \rightarrow \infty$ [37]. However, the giant cluster emerges at $\langle k \rangle = 1$ and its size monotonically increases as $\langle k \rangle$ increases with the order parameter (the relative size of the giant cluster) critical exponent $\beta = 1$. The giant cluster size G follow $G \sim (t - t_c)^\beta$ at the transition point $t_c = 1/2$. Moreover, the susceptibility, which is the average cluster size, diverges at $t = t_c$ and the corresponding exponents $\gamma = 1$. The exponents τ and σ for the cluster size distributions are 2.5 and 0.5, respectively.

1.3 Suppression effects

We consider the local and global suppression effects. Under the local suppression rules, the candidate nodes are selected with the limited information. The examples of local suppression rules are minimal [26] and da Costa [19, 38–40] rules. On the other hand, the global suppression rule needs the information of all nodes in the system to select candidate nodes. we consider the global suppression rule following the half-restricted process [41] in static [42] and growing [27, 28] networks.

1.4 Growing simplicial complexes

There are more than two-body interactions in the real world, and these multi-body interactions [43–45] can be explained by simplexes in simplicial complexes [46, 47]. In this framework, considering that networks consist of 0-

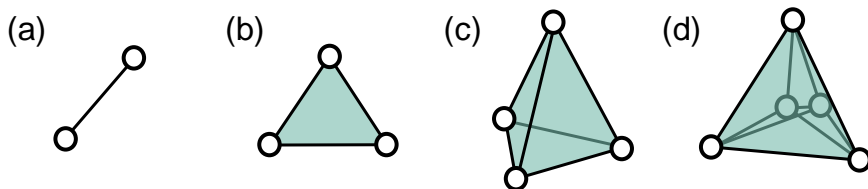


Fig. 1.1: (Color online) Illustrations of d -simplexes for $d =$ (a) 1, (b) 2, (c) 3, and (d) 4. In each figure, 0-simplex is represented by an open circle.

simplexes for nodes and 1-simplexes for edges, we extend the growing networks to the growing simplicial complexes where a 0-simplex is added to the system and a d -simplex is added with probability p every time step. When $d + 1$ number of 0-simplexes are selected to compose a new d -simplex, each 0-simplex has an attractiveness of their facet degree with an initial attractiveness a , leading to the scale-free structure. Illustrations of d -simplexes are presented in Fig. 1.1

Chapter 2

Local suppression effects in growing networks

2.1 Minimal rule

2.1.1 Model

We consider two types of network models, growing and static. In a growing network, the number of nodes increases one by one at each time step, whereas in a static model, the number of nodes remains fixed from the beginning. Links are added one by one at each time step in both models according to the following rules:

(i) A growing network begins with isolated nodes in a system. At each time step, a node is added in the system and then m candidate nodes are selected randomly. At time t , when the number of nodes $N(t) = 1 + t$ is less than m , all of the nodes are selected as candidates. When $N(t) \geq m$, m clusters to which the respective m nodes belong are identified. Some of those clusters may be identical when they contain more than one selected node. The two smallest clusters among those m clusters are selected, and the two corresponding nodes are identified and are connected with probability p if they are not already connected. When $m = 2$, this growing network model reduces to the exponentially growing network model, which was proposed

by Callaway et al. [14].

(ii) For the static network model, N nodes are present from the beginning and remain fixed. At each time step, m candidate nodes are selected uniformly at random, and the sizes of the respective clusters where they belong are identified. The two nodes corresponding to the two smallest clusters are connected with probability one. When $m = 2$, this static network model reduces to the Erdős-Rényi (ER) random network model [36].

2.1.2 Rate equation approach for the cluster size distribution

Growing network model with $m = 3$

Let $n_s(p, t)$ be the number of clusters of size s divided by $N(t)$, where p denotes the probability that a links is connected between two selected nodes.

The rate equation of $n_s(p, t)$ is given by

$$\begin{aligned} \frac{d(N(t)n_s)}{dt} = p \left[\sum_{i+j=s; i < j} 3in_i(jn_j)^2 + \sum_{i+j=s; i < j} 6in_ijn_jc_{j+1} + \left(\frac{s}{2}n_{\frac{s}{2}}\right)^3 \right. \\ \left. + 3\left(\frac{s}{2}n_{\frac{s}{2}}\right)^2 (c_{\frac{s}{2}+1}) - 2(sn_s)^3 - 6(sn_s)^2c_{s+1} - 3(sn_s)^2(1-c_s) \right. \\ \left. - 3sn_s(c_{s+1})^2 - 6sn_s(1-c_s)c_{s+1} \right] + \delta_{1s}, \end{aligned} \quad (2.1)$$

where n_i denotes $n_i(p, t)$ and $c_s(p, t) = 1 - \sum_{i < s} in_i(p, t)$ to simplify the notation. The first term, $3in_i(jn_j)^2$, of the right hand side of Eq. (2.1) comes from the merging of two clusters of size i and j with $i < j$, which produces a cluster of size $s = i + j$. One node is selected from a cluster of size i

and the other two nodes are selected from either i) one cluster or ii) two distinct clusters of the size j . However, the probability to occur the case i) is $jn_i(j/N)$, which is much smaller than $(jn_j)^2$ for the case ii), and thus the first case was ignored. The factor 3 comes from the combinatorics of the three possible clusters. For simplicity, this process is denoted as $(i, j > i, j > i)_{i+j=s}$. The second term comes from the process $(i, j > i, k > j)_{i+j=s}$. This means that the three nodes are chosen from the clusters of different sizes i, j and k with the constraint $i < j < k$, and the two smallest clusters of sizes i and j are merged. Since the size k can be arbitrary as long as $k > j$, we used $c_{j+1} \equiv \sum_{k>j} kn_k$ instead of kn_k . The factor 6 again comes from the combinatorics. The third term comes from the process $(\frac{s}{2}, \frac{s}{2}, \frac{s}{2})$, which represents that three nodes are chosen from three distinct clusters of size $s/2$. Similarly, the fourth term comes from the process $(\frac{s}{2}, \frac{s}{2}, i > \frac{s}{2})$ which means that two nodes are chosen from two distinct clusters of size $s/2$ and the other node are chosen from a cluster of size k larger than $s/2$. Note that if two nodes are chosen from the same cluster of size $s/2$, then there is no cluster merging and the probability to occur this case becomes $(\frac{s}{2}n_{s/2})(s/2)(1/N)$, which is smaller than $(\frac{s}{2}n_{s/2})^2$. Thus, this case was ignored. The factor 3 again comes from the combinatorics. The third and fourth terms appear only when s is even. Up to here all terms are for cluster creations.

The terms from the fifth to the ninth are for the annihilation of clusters of size s . The fifth and the sixth terms come from the processes (s, s, s) and $(s, s, i > s)$, respectively. These two terms correspond to the third and the fourth terms before. But the prefactors are two times bigger because two clusters of size s merge and annihilate simultaneously. The seventh term

comes from the process $(s, s, i < s)$ which represents that two nodes are chosen from two distinct clusters of size s and the other one node is chosen from a cluster of size i smaller than s . Again we remark that we ignore the case that two or more nodes are selected from the same cluster, because that case occurs with smaller probability than that of two or more nodes are selected from individually distinct clusters in the limit $N(t) \rightarrow \infty$. The factor 3 again comes from the combinatorics. The eighth term comes from the process $(s, i > s, i > s)$, which represents the case that one node is chosen from a cluster of size s and the other two nodes are chosen from the clusters of size i larger than s . The factor 3 again comes from the combinatorics. The ninth term comes from the process $(s, i < s, j > s)$ which means that three nodes are chosen from three different clusters of size s , smaller than s and bigger than s , respectively. The factor 6 comes from the combinatorics. Overall factor p is the probability that the determined two nodes are connected. The last term δ_{1s} arises when a node is added every time step.

Based on this rate equation, we calculate $n_s(p)$ in the steady state up to a certain size s^* , for instance, $s^* = 10^6$. Note that $n_s(p)$ decays in a power-law way as $n_s(p_c) \sim s^{-\tau}$ at a transition point p_c , and exhibits crossover behavior $n_s(p) \sim s^{-\tau} \exp(-s/s_c)$ for $p \neq p_c$ with $s_c \sim |p - p_c|^{-1/\sigma}$ [1, 2]. When $p > p_c$, an infinite cluster exists separately from the finite clusters. The percolation threshold is calculated as $p_c = 0.413842(1)$ using the criterion that $n_s(p_c)$ follows power law at p_c as shown in Fig. 2.2. Moreover, the exponent τ is determined to be $\tau \approx 2.5$. We also check the crossover behaviors for $p < p_c$ and $p > p_c$ in Fig. 2.2. The exponent σ is obtained by scaling the plots of $n_s(p)s^\tau$ versus $s|p - p_c|^{1/\sigma}$ for different p values. Fig. 2.3 can

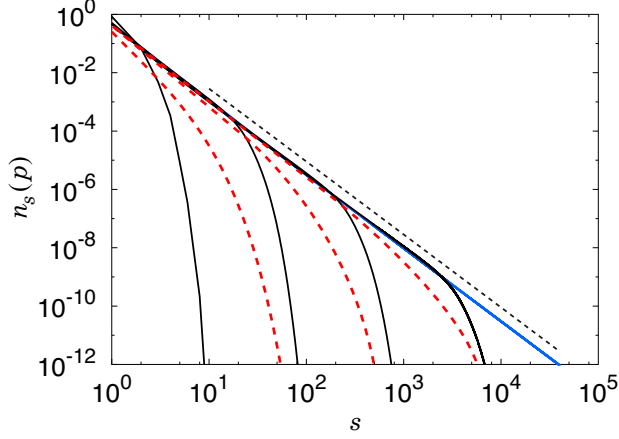


Fig. 2.2: (Color online) For growing networks with $m = 3$, plot of $n_s(p)$ vs s at $p = p_c$ (blue solid line), $p > p_c$ (red dashed curves) and $p < p_c$ (black solid curves) based on numerical values obtained from the rate equation. The transition point p_c is $p_c = 0.413842(1)$, and the exponent τ is approximately 2.5. The black dashed line is a guide line with slope -2.5 .

be shown that the data are well collapsed on a single curve when $\sigma \approx 0.72$.

Next, the order parameter is obtained using the relation, $G(p) \approx 1 -$

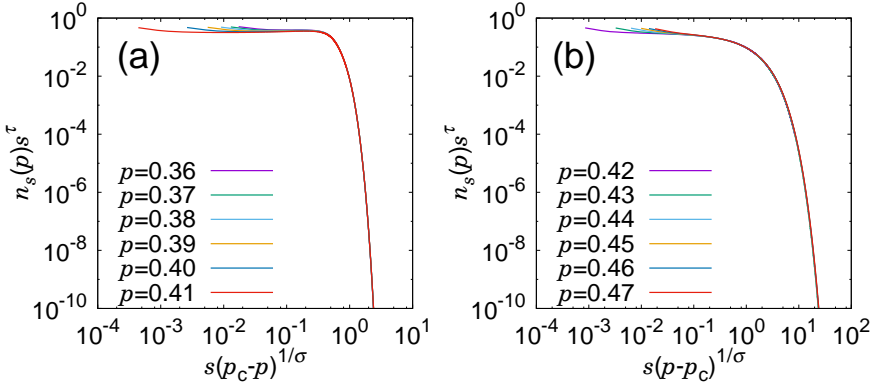


Fig. 2.3: (Color online) For growing networks with $m = 3$, plot of $n_s(p)s^\tau$ versus $s|p - p_c|^{1/\sigma}$ for different values of p when (a) $p < p_c$ and (b) $p > p_c$. Data for different p values are well collapsed onto a single curve by choosing $\sigma = 0.720(2)$ and $\tau = 2.500(1)$.

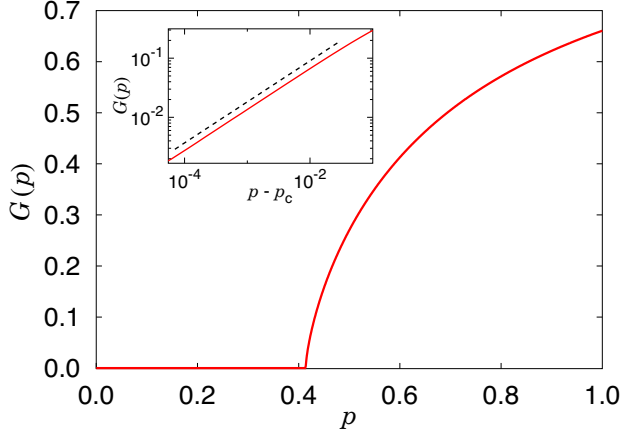


Fig. 2.4: (Color online) For growing networks with $m = 3$, plot of $G(p)$ vs p . The data points are obtained from the rate equation. Inset: The dashed line is a guide line with slope $0.694(2)$.

$\sum_{s=1}^{s^*} s n_s(p)$ [1, 2], where s^* takes on several values to observe the effect of the artificially established cutoff values. The order parameter follows the power-law form, $G(p) \sim (p - p_c)^\beta$, where $\beta = 0.694(2)$. The inset of Fig. 2.4 is a double logarithmic plot of the order parameter as a function of $(p - p_c)$, which exhibits power-law behavior as expected. The obtained value of β satisfies the hyperscaling relation $\beta = (\tau - 2)/\sigma$ [1, 2].

The mean cluster size $\langle s \rangle$ is obtained from the cluster size distribution as $\langle s \rangle = \sum_{s=1}^{s^*} s^2 n_s(p)$, which behaves like the susceptibility, $\langle s \rangle \sim (p - p_c)^{-\gamma}$ for $p > p_c$ and $(p_c - p)^{-\gamma'}$ for $p < p_c$. We also determine that $\gamma = \gamma' \approx 0.696$. The numerical values obtained from the rate equation are shown in Fig. 2.5. In the insets, $\langle s \rangle$ is plotted in double logarithmic axes as a function of $p - p_c$ for $p > p_c$, and $p_c - p$ for $p < p_c$. The exponent γ satisfies the well-known scaling relation $\gamma = (3 - \tau)/\sigma$ [1, 2].

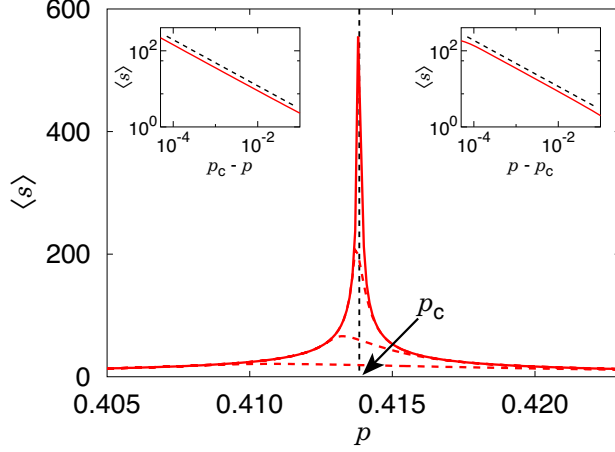


Fig. 2.5: (Color online) For growing model with $m = 3$, plot of the susceptibility, that is mean cluster size, as a function of p . The data points are obtained from the rate equation. Insets : Double logarithmic plots of $\langle s \rangle$ versus $|p - p_c|$ for $p < p_c$ (left) and $p > p_c$ (right). The dashed lines are guide lines with slope $-0.696(3)$.

Growing model with general m

We extend the rate equation in Eq. (2.1) for $m = 3$ to arbitrary m as follows:

$$\begin{aligned}
 \frac{d(N(t)n_s)}{dt} = & p \left[\sum_{r=1}^{m-1} m \binom{m-1}{r-1} \sum_{i+j=s; i < j} in_i(jn_j)^{m-r} (c_{j+1})^{r-1} \right. \\
 & + \sum_{r=1}^{m-1} \binom{m}{r-1} \left(\frac{s}{2} n_{\frac{s}{2}} \right)^{m-(r-1)} (c_{\frac{s}{2}+1})^{r-1} \\
 & - 2 \sum_{r=2}^m \binom{m}{r} (sn_s)^r (c_{s+1})^{m-r} - m (sn_s) (c_{s+1})^{m-1} \\
 & \left. - \sum_{r=1}^{m-1} m \binom{m-1}{r} (1 - c_s) (sn_s)^r (c_{s+1})^{m-1-r} \right] + \delta_{1s}. \quad (2.2)
 \end{aligned}$$

Again, the second term on the right hand side is valid only when s is even.

Repeating the steps taken in the case $m = 3$, we obtain the critical exponents

τ , σ , β , γ , and the percolation threshold p_c up to $m = 10$, which are listed

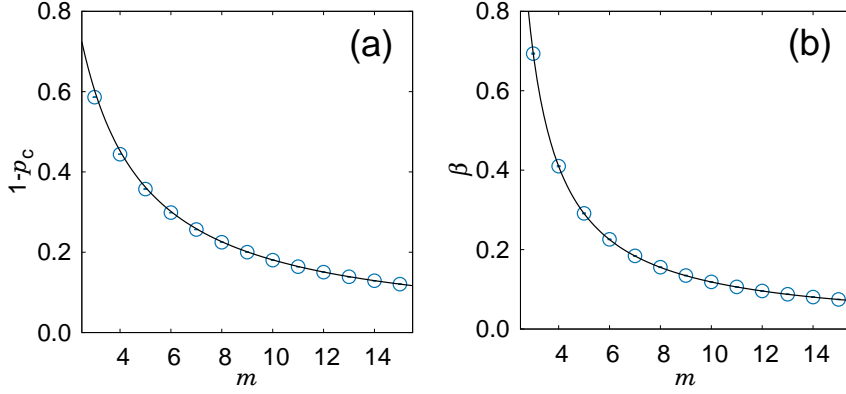


Fig. 2.6: (Color online) For growing model with general m , (a) empirical plot of $1 - p_c$ versus m . Data points of $1 - p_c$ for different values of m behave like the formula $1.81/m$. (b) Empirical plot of β vs m . Data points of β for different values of m behave like the formula $\beta = 1/(m - 1.56)$. Note that the error bars are smaller than the symbol sizes.

in Table 2.1.

Following the conventional formalism for the second-order percolation transition, we examine the scaling relation between the critical exponents and their tendencies for m candidates. Note that the critical point p_c and critical exponent β behave like $1 - p_c \approx 1.81/m$ and $\beta \approx 1/(m - 1.56)$ as shown in Fig. 2.6. However, a rigorous derivation of these formulas is still necessary. Next, we determine the exponents τ and σ for $m = 4 \dots 10$ by following similar steps used for $m = 3$. We find that the values are approximately by the formulae $\tau = 2 + 1/(m - 1)$ and $1/\sigma = (m - 1)\beta$ as shown in Fig. 2.24. Furthermore, we determine that $\gamma = (m - 2)\beta$.

Table. 2.1: Numerical estimates of the percolation threshold p_c , exponent of the cluster size distribution τ , exponent of the characteristic cluster size σ , exponent of the order parameter β , and exponent of the susceptibility γ of the growing network model for $m = 3, \dots, 10$. τ^* and β^* were obtained from $\tau^* = 2 + 1/(m - 1)$ and $\beta^* \approx 1/(m - 1.56)$, respectively.

m	p_c	τ^*	τ	σ	β^*	β	γ
3	0.413842(1)	$\frac{5}{2}$	2.500(1)	0.720(2)	0.694	0.694(2)	0.696(3)
4	0.555873(1)	$\frac{7}{3}$	2.333(1)	0.812(2)	0.410	0.410(2)	0.813(3)
5	0.642748(1)	$\frac{9}{4}$	2.250(1)	0.858(2)	0.291	0.291(1)	0.874(6)
6	0.701282(1)	$\frac{11}{5}$	2.200(1)	0.885(2)	0.225	0.226(1)	0.904(2)
7	0.743370(1)	$\frac{13}{6}$	2.167(1)	0.905(2)	0.184	0.184(1)	0.922(2)
8	0.775078(1)	$\frac{15}{7}$	2.143(1)	0.918(2)	0.155	0.156(1)	0.934(2)
9	0.799820(1)	$\frac{17}{8}$	2.125(1)	0.928(2)	0.134	0.135(1)	0.944(3)
10	0.819663(1)	$\frac{19}{9}$	2.111(1)	0.936(2)	0.119	0.119(1)	0.950(3)

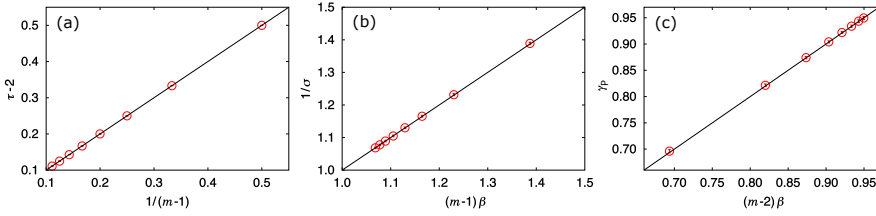


Fig. 2.7: (Color online) For growing networks with general m , formula testing for the exponents of (a) the cluster size distribution τ , (b) the characteristic cluster size σ , and (c) the mean cluster size γ , where the numerical data are obtained from the rate equation. Data are fit reasonably to the straight line predicted by the formula, and the error bars are smaller than the symbol sizes.

Static model with $m = 3$

We consider the evolution of static networks under the rule described at the first in Sec. 2.1. In this case, the number of nodes is fixed all the way as N .

The rate equation is written as

$$\begin{aligned}
N \frac{dn_s}{dt} = & \sum_{i+j=s; i < j} 3in_i(jn_j)^2 + \sum_{i+j=s; i < j} 6in_i jn_j c_{j+1} + \left(\frac{s}{2} n_{\frac{s}{2}}\right)^3 \\
& + 3 \left(\frac{s}{2} n_{\frac{s}{2}}\right)^2 \left(s_{\frac{s}{2}+1}\right) - 2(sn_s)^3 - 6(sn_s)^2 c_{s+1} - 3(sn_s)^2 (1 - c_s) \\
& - 3sn_s(c_{s+1})^2 - 6sn_s(1 - c_s)c_{s+1},
\end{aligned} \tag{2.3}$$

where n_i denotes $n_i(t)$ and $c_s(t) = 1 - \sum_{i < s} in_i(t)$. The terms on the second line of Eq. (2.3) related with $\frac{s}{2}$ are valid only when s is even. In contrast to the growing network, there is no steady state in the size distribution, and n_s depends on t . Accordingly it takes longer time to evaluate $n_s(t)$ explicitly compared with that of the growing network model. We obtain $n_s(t)$ up to a certain cluster size $s^* = 5 \times 10^5$.

We determine the percolation threshold t_c as shown in Fig. 2.8 by the criterion that the cluster size distribution follows power law at t_c . It is obtained that $t_c = 0.849130(1)$ and $n_s(t_c) \sim s^{-\tau}$ with $\tau \approx 2.105$. For $t < t_c$ and $t > t_c$, the cluster size distribution exhibits a crossover behavior as $n_s(t) \sim s^{-\tau} \exp(-s|t - t_c|^{1/\sigma})$. Using the data-collapse method, we obtain $\sigma \approx 0.79$ as shown in Fig. 2.9.

Next, we consider the behavior of the order parameter $G(t)$ at time step t . The order parameter is calculated using the relation $G(t) = 1 - \sum_{s=1}^{s^*} sn_s(t)$. We expect that $G(t) \sim (t - t_c)^\beta$, and obtain $\beta = 0.133(1)$ in Fig. 2.10. We also obtain the mean cluster size or the susceptibility defined as $\langle s \rangle = \sum_{s=1}^{s^*} s^2 n_s(t)$. Following the convention, it behaves as $\langle s \rangle \sim |t - t_c|^{-\gamma}$. We estimate that $\gamma = 1.131(6)$ in Fig. 2.11. The obtained exponent values $\beta =$

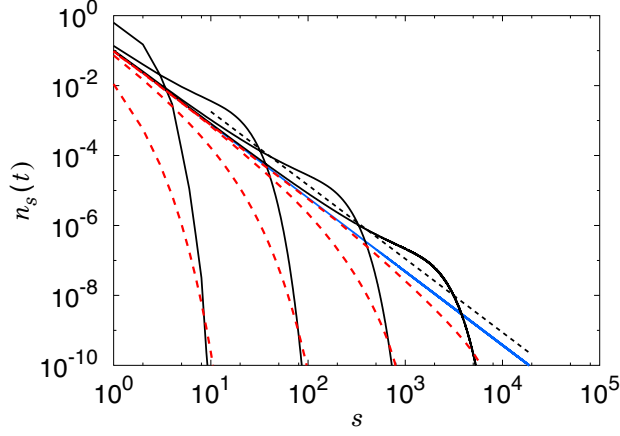


Fig. 2.8: (Color online) For static networks with $m = 3$, plot of $n_s(t)$ vs s at $t = t_c$ (blue solid line), $t > t_c$ (red dashed curves) and $t < t_c$ (black solid curves) based on the numerical values obtained from the rate equation. The transition point t_c is determined as $t_c = 0.849130(1)$ and the exponent τ is determined as $\tau = 2.105(5)$. Black dashed line is a guideline with slope -2.105 .

0.133 and $\gamma = 1.133$ satisfy the scaling relation $\beta = (\tau - 2)/\sigma$ and $\gamma = (3 - \tau)/\sigma$, respectively.

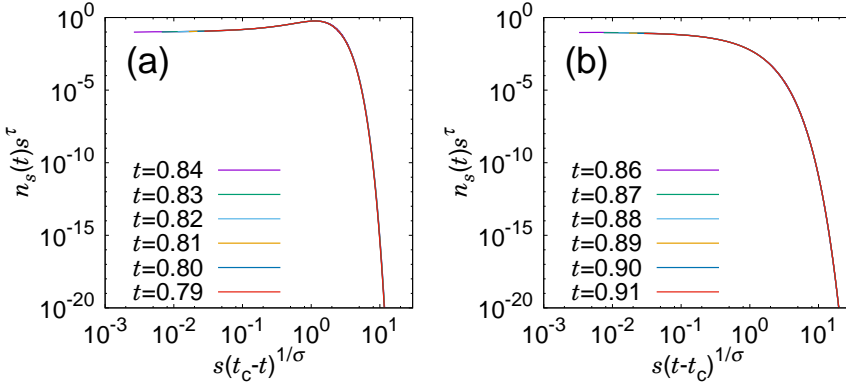


Fig. 2.9: (Color online) For static networks with $m = 3$, scaling plot of $n_s(t)s^\tau$ versus $s|t - t_c|^{1/\sigma}$ for different t that are (a) less and (b) greater than t_c . Taking $\tau = 2.105(5)$ and $\sigma = 0.790(1)$, the data for different t values look collapsed onto a single curve.

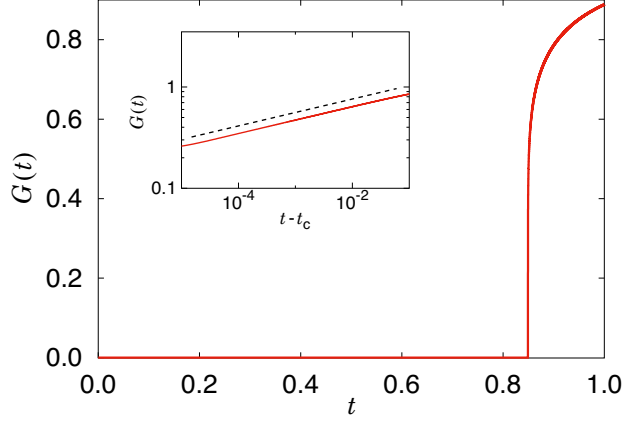


Fig. 2.10: (Color online) For static networks with $m = 3$, plot of the order parameter $G(t)$ as a function of t . Inset : The dashed line is a guide line with slope $0.133(1)$.

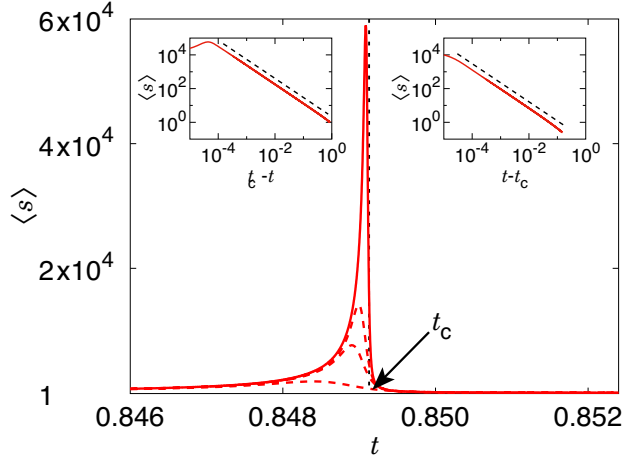


Fig. 2.11: (Color online) For static networks with $m = 3$, plot of $\langle s \rangle$ as a function of t . Inset : Plot of the susceptibility, the mean cluster size as a function of t for $t > t_c$ (right) and $t < t_c$ (left). The dashed lines are guide lines with slope $-1.131(6)$.

Table. 2.2: Numerical estimates of the percolation threshold t_c , exponent of the cluster size distribution τ , exponent of the characteristic cluster size σ , exponent of the order parameter β , and exponent of the susceptibility γ of the static network model for $m = 2, \dots, 5$. Subsequently, τ^* and β^* were obtained from $\tau^* = 2 + \beta/[1 + (m-1)\beta]$ and $\beta^* \approx 0.465 \exp(-0.70m)$, respectively.

m	t_c	τ^*	τ	σ	β^*	β	γ
2	0.5	2.5	2.5	0.5	0.115	1	1
3	0.849130(1)	2.105	2.105(1)	0.790(1)	0.057	0.133(1)	1.131(6)
4	0.939678(1)	2.037	2.037(1)	0.890(1)	0.028	0.042(1)	1.082(6)
5	0.972672(1)	2.016	2.015(2)	0.940(1)	0.014	0.017(1)	1.050(4)

Static network model with general m

We extend the rate equation for $m = 3$ to an arbitrary value of m as follows:

$$\begin{aligned}
N \frac{dn_s}{dt} = & \sum_{r=1}^{m-1} m \binom{m-1}{r-1} \sum_{i+j=s; i < j} in_i (jn_j)^{m-r} (c_{j+1})^{r-1} \\
& + \sum_{r=1}^{m-1} \binom{m}{r-1} \left(\frac{s}{2} n_{\frac{s}{2}} \right)^{m-(r-1)} (c_{\frac{s}{2}+1})^{r-1} \\
& - 2 \sum_{r=2}^m \binom{m}{r} (sn_s)^r (c_{s+1})^{m-r} - m (sn_s) (c_{s+1})^{m-1} \\
& - \sum_{r=1}^{m-1} m \binom{m-1}{r} (1-c_s) (sn_s)^r (c_{s+1})^{m-1-r}, \tag{2.4}
\end{aligned}$$

where the second term of the right hand side is valid only when s is an even number.

Taking similar steps used for $m = 3$, we determine the transition points and critical exponent β for general m up to $m = 15$. We determine empirically that these values behave asymptotically like $1 - p_c \approx \exp(-0.59m)$ and $\beta \approx \exp(-0.70m)$, respectively. This conjecture was alluded to in [38, 39]. A numerical test is shown in Fig. 2.12.

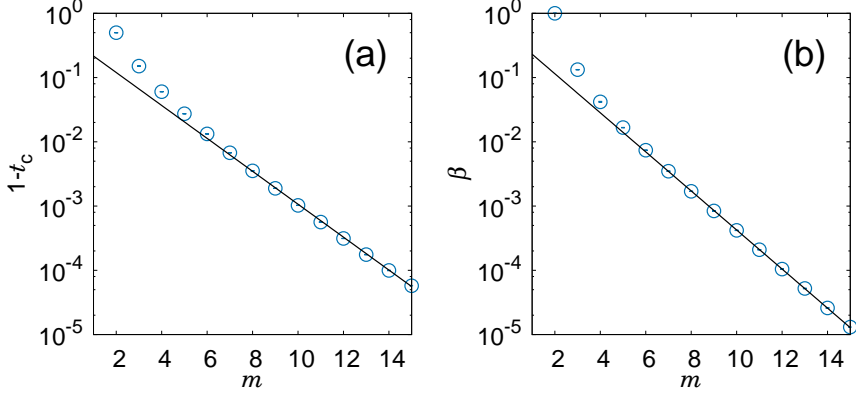


Fig. 2.12: (Color online) For static networks, (a) plot of $1 - t_c$ versus m on a semi-logarithmic scale. (b) Plot of the estimated values of the exponent β for general m versus m on a semi-logarithmic scale. Asymptotically, the data points likely lie on a straight line. The error bars are smaller than the symbol sizes.

Furthermore, we determine the exponent values τ and σ for $m = 4$ and $m = 5$ because of the instability of the cluster size distribution when the exponents τ and σ are calculated in the vicinity of $\tau = 2$. The obtained values are listed in Table 2.2. Notice that the values approximate the formulas $\tau = 2 + \beta/[1 + (m - 1)\beta]$ and $1/\sigma = 1 + (m - 1)\beta$, as shown in Fig. 2.13. This conclusion is based on a previous analytic solution to the model in [38, 39]. Due to a slight difference in the dynamic rule, the value

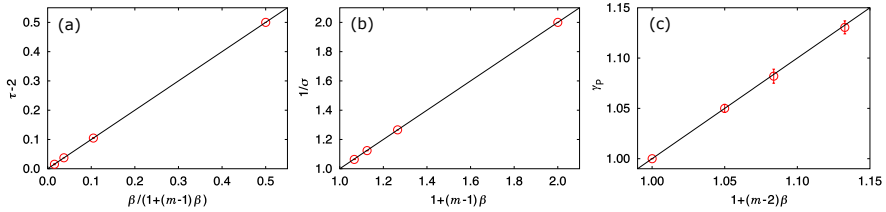


Fig. 2.13: (Color online) For static networks, formula testing for the exponents of (a) the cluster size distribution τ , (b) the characteristic cluster size σ , and (c) the mean cluster size γ , where the numerical data is obtained from the rate equation. Note that the data fit reasonably to the straight line predicted by the formula.

of m in our model corresponds to $2m$ in [38, 39]; thus, the analytic solution in [38, 39] is valid for our model by replacing $2m$ with m . This allows us to obtain $\gamma = 1 + (m - 2)\beta$.

2.1.3 Monte Carlo simulations

Growing network model with $m = 3$

To determine the exponents τ and σ for the cluster size distribution, we numerically perform Monte Carlo simulations of the growing network models for different system sizes $N/10^4 = 2^3 - 2^{10}$; the ensemble average is taken over 10^4 configurations.

We first examine the cluster size distribution for several values of p around the transition point p_c in Fig. 2.14. The cluster size distribution follows power law at p_c and exhibits crossover behavior of $n_s(p) \sim s^{-\tau} \exp(-s|p - p_c|^{1/\sigma})$. We determine $p_c = 0.4138(2)$ using the criteria that at p_c ,

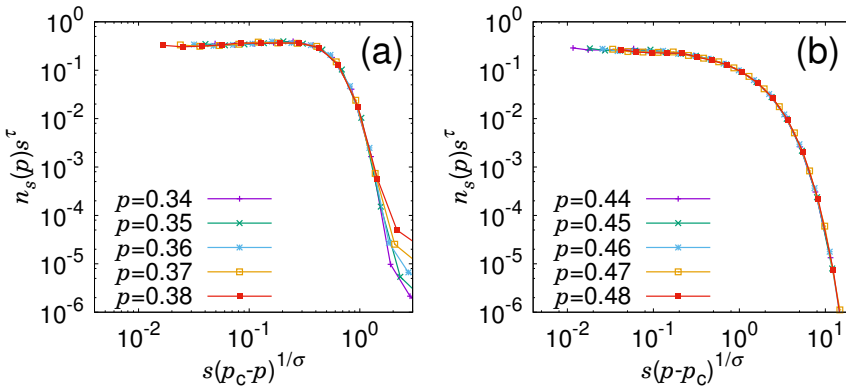


Fig. 2.14: (Color online) To obtain τ and σ by Monte Carlo simulations for growing networks, scaling plot of $n_s(p)s^\tau$ versus $s|p - p_c|^{1/\sigma}$. Data are collapsed onto a single curve by choosing $\tau = 2.5$ and $\sigma = 0.72$ for (a) $p < p_c$ and (b) $p > p_c$.

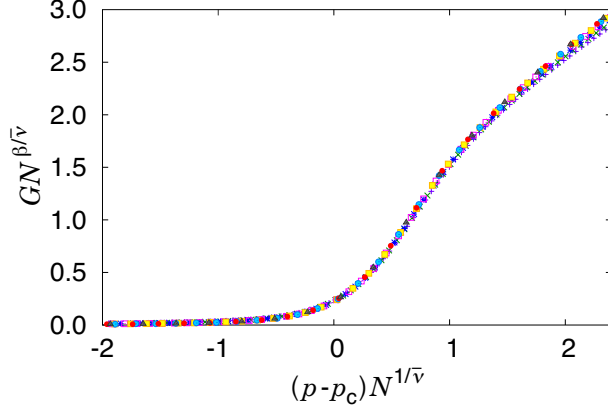


Fig. 2.15: (Color online) To obtain β by Monte Carlo simulations for growing networks, scaling plot of $GN^{\beta/\bar{\nu}}$ versus $(p - p_c)N^{1/\bar{\nu}}$ for system sizes $N/10^4 = 2^3 - 2^{10}$. Data are collapsed onto a single curve with the values of $1/\bar{\nu} = 0.35(3)$ and $\beta/\bar{\nu} = 0.24(3)$.

$n_s(p_c)$ decays in a power-law way and the relative size of the largest cluster, $G_N(p)$, follows a power law, $G \sim N^{-\beta/\bar{\nu}}$. Using the data-collapse method,

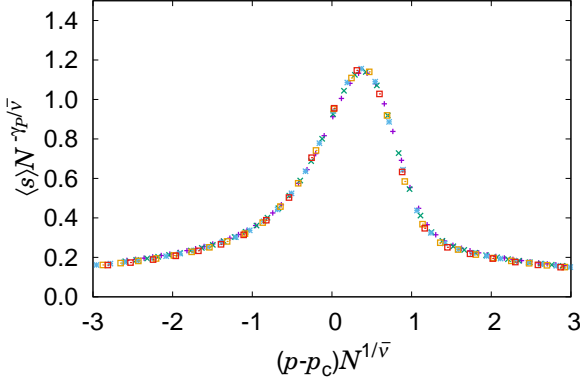


Fig. 2.16: (Color online) To obtain the exponent γ using Monte Carlo simulations for growing networks, data collapse plot of $\langle s \rangle N^{-\gamma/\bar{\nu}}$ versus $(p - p_c)N^{1/\bar{\nu}}$ for the system sizes $N/10^4 = 2^6 - 2^{10}$ in growing networks. The exponent values are $\gamma = 0.696$ and $1/\bar{\nu} = 0.35$.

we determine the exponent values of τ and σ to be $\tau \approx 2.5$ and $\sigma \approx 0.72$, respectively, which are in good agreement with the values obtained by the rate equation approach.

By measuring the exponent of the power-law behavior of $G_N(p)$ and using a finite-size scaling formula $G_N(p) = N^{-\beta/\bar{\nu}} f((p - p_c)N^{1/\bar{\nu}})$, we determine the ratios $\beta/\bar{\nu} = 0.24(3)$ and $1/\bar{\nu} \approx 0.35(3)$, as shown in Fig 2.15. We determine an exponent value of $\beta \approx 0.69$. These values are consistent with those obtained from the rate equations.

The susceptibility is also examined by plotting it in scaling form, i.e., $\langle s \rangle N^{-\gamma/\bar{\nu}}$ versus $(p - p_c)N^{1/\bar{\nu}}$ with $\gamma = 0.696$ and $1/\bar{\nu} = 0.35$ in Fig. 2.16 for different sizes $N/10^4 = 2^6 - 2^{10}$; the ensemble average is taken over 10^4 configurations. Notice that the data are well collapsed. This means that the hyperscaling relation $\bar{\nu} = 2\beta + \gamma$ does not hold.

Static network model with $m = 3$

To determine the exponents τ and σ for the cluster size distribution, we numerically perform Monte Carlo simulations of the static network models for different system sizes $N/10^4 = 2^0 - 2^{10}$. The ensemble average is taken over 10^5 for each data point. The cluster size distributions $n_s(t)$ for different times are plotted in scaling form, i.e., $n_s(t) \sim s^{-\tau} \exp(-s|t - t_c|^{1/\sigma})$, as shown in Fig. 2.17. Using the previously obtained values $t_c = 0.84913(1)$ and $\tau \approx 2.1$, we determine that the data for different t are well collapsed onto a single curve with $\sigma \approx 0.79$.

Next, we consider the order parameter $G(t)$ as a function of the time step t for different sizes $N/10^4 = 2^0 - 2^{10}$. The critical point t_c and critical

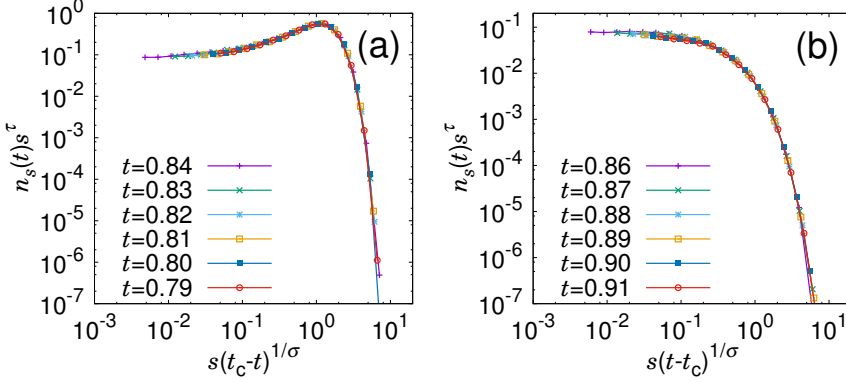


Fig. 2.17: (Color online) To obtain the exponents τ and σ using Monte Carlo simulations for static networks, data collapse plots of the rescaled cluster size distribution $n_s(t)s^\tau$ versus $s|t - t_c|^{1/\sigma}$ for different time steps when (a) $t < t_c$ and (b) $t > t_c$, where $\tau = 2.105$ and $\sigma = 0.79$.

exponent β are determined using the scaling ansatz $G_N(t) = N^{-\beta/\bar{\nu}} f((t - t_c)N^{1/\bar{\nu}})$. Using the criterion that $G_N(t) \sim N^{-\beta/\bar{\nu}}$ at $t = t_c$, we determine $t_c = 0.84913(1)$ and $\beta/\bar{\nu} \approx 0.06$ in Fig. 2.18. Moreover, all of the data for different system sizes are systematically collapsed onto a single curve when $1/\bar{\nu} \approx 0.45$, as shown in Fig. 2.18. This suggests that $\beta \approx 0.133$. The value of β is consistent with the results obtained by the rate equation approach.

Finally, we study the susceptibility behavior as a function of the time step. The susceptibility is also examined by plotting it in scaling form, i.e., $\langle s \rangle N^{-\gamma/\bar{\nu}}$ versus $(t - t_c)N^{1/\bar{\nu}}$ with $\gamma = 1.133$ and $1/\bar{\nu} = 0.45$ in Fig. 2.19. Notice that the data are well collapsed. This means that the hyperscaling relation $\bar{\nu} = 2\beta + \gamma$ does not hold.

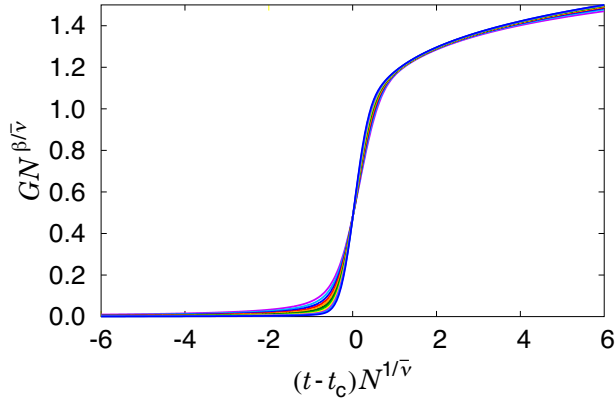


Fig. 2.18: (Color online) To obtain β using Monte Carlo simulations for static networks, data collapse plot of $GN^{\beta/\bar{\nu}}$ versus $(t-t_c)N^{1/\bar{\nu}}$ for system sizes $N/10^4 = 2^0 - 2^{10}$. Data for different values of N are systematically collapsed near the transition point by taking $1/\bar{\nu} = 0.45$ and $\beta/\bar{\nu} = 0.06$.

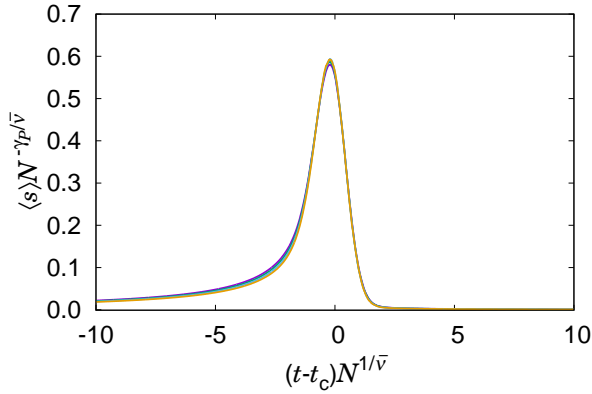


Fig. 2.19: (Color online) To obtain γ using Monte Carlo simulations for static networks, data collapse plot of $\langle s \rangle N^{-\gamma/\bar{\nu}}$ versus $(p-p_c)N^{1/\bar{\nu}}$ for system sizes $N/10^4 = 2^7 - 2^{10}$, where the exponent values $\gamma = 1.133$ and $1/\bar{\nu} = 0.45$ are used.

2.1.4 Comparison of $n_s(p)$ for growing network models

It is interesting to note that the percolation occurring in growing network models when $m \geq 3$ is a second-order phase transition, whereas it is of infinite order when $m = 2$. We investigate the cluster size distribution $n_s(p)$ for $m = 2$. As shown in Fig. 2.20, $n_s(p)$ decays in a power-law way when $p \leq p_c$, while it exhibits crossover behavior when $p > p_c$. The power-law behavior of $n_s(p)$ when $p \leq p_c$ implies that the region $p \leq p_c$ is the critical phase, which is noticeable in the infinite-order transition. Intuitively, when $p \leq p_c$, in the growing network, the fraction of nodes that belong to small-size clusters is relatively low compared to the fraction for the second-order phase transition model; for example, when $m \geq 3$ in Fig. 2.2. However, when

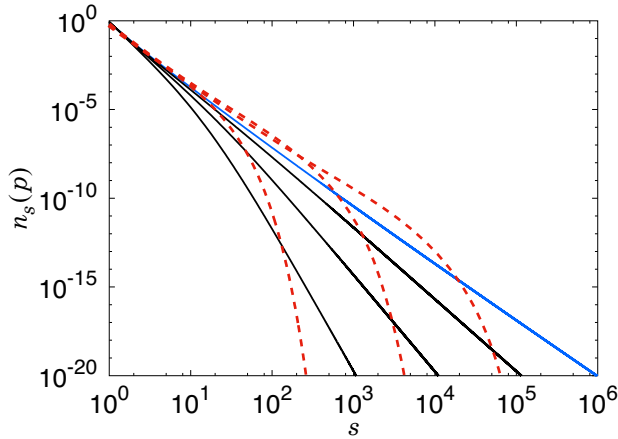


Fig. 2.20: (Color online) For growing network with $m = 2$, plot of $n_s(p)$ versus s at $p = p_c$ (blue solid line), $p > p_c$ (red dashed line), and $p < p_c$ (black solid line) based on the numerical values obtained from the rate equation. The transition point is $p_c = 0.125$. For $p \leq p_c$, $n_s(p)$ decays in a power-law manner, indicating that the transition is infinite-order.

$m \geq 3$, the density of large-size clusters is suppressed by the Achlioptas rule, which leads to crossover behavior even for $p < p_c$. Thus, the percolation transition in growing networks for $m \geq 3$ is second order.

2.1.5 Analysis

In our minimal model that incorporates both growing and static Achlioptas processes, the results obtained from the rate equations and Monte-Carlo simulations for the cluster size distribution are consistent. When $m = 2$, the growing and static models correspond to the Callaway random growing model [14] and Erdős-Rényi model [36], respectively.

In the growing network model, as m increases from 2 to 3, the transitional nature of percolation changes from infinite-order to second-order due to the Achlioptas process [16]. On the other hand, in the static model, the order of the phase transition is the same as that of the second-order ER model, but the order parameter exponent β decreases exponentially as m increases and the transition becomes more explosive. The Achlioptas process rule leads to the suppression effect against the growth of large clusters, which causes the cluster size distribution in large-cluster regions to decay exponentially; thus, the transition is second-order.

Moreover, in this paper, we showed that the critical exponent β decreases algebraically with m in growing networks; however, it decays exponentially in static networks. This fact reflects that the suppression effect in growing networks is weaker than that in static networks. Furthermore, we obtained the critical exponents and their tendencies in both growing and static models for arbitrary values of m . We also found that the m -

dependent exponents always satisfy the scaling relations $\beta = (2 - \tau)/\sigma$ and $\gamma = (3 - \tau)/\sigma$ [1, 2]. However, the hyperscaling relation $\bar{\nu} = 2\beta + \gamma$ does not hold in both growing and static networks. We expected $\bar{\nu} \approx 2.08$ from the relation $2\beta + \gamma$ for the growing networks of $m = 3$, but obtained $\bar{\nu} \approx 2.86$ ($1/\bar{\nu} = 0.35$) from Monte Carlo simulations. For the static network with $m = 3$, we expected $\bar{\nu} \approx 1.4$ ($1/\bar{\nu} = 0.71$), but obtained $\bar{\nu} = 2.22$ ($1/\bar{\nu} = 0.45$). The origin of these inconsistencies are still not clear. We remark that the failure of the hyperscaling relation $\bar{\nu} = 2\beta + \gamma$ was also observed in the previous research of Achlioptas process [22].

2.2 da Costa rule

2.2.1 Model and rate equation

We consider a growing network model under a rule that suppresses the growth of large clusters locally with limited information. It consists initially of an isolated node, and a new node is added to the system at each time step; consequently, the total number of nodes at time t is $N(t) = t + 1$. Then, two sets of m candidate nodes are selected randomly. The node that belongs to the smallest cluster in each set is selected, and these two nodes are connected with the wiring probability p , as depicted schematically in Fig. 2.21. When $m = 1$, this growing network model reduces to the GRN model proposed by Callaway et al. [14]. This type of suppression rule in static network models was first considered by da Costa et al. [19]. When $m = 1$, this da Costa model also becomes the ER random network model [36]. The similar, but simpler, minimal rule is considered for the growing and static network models [26]

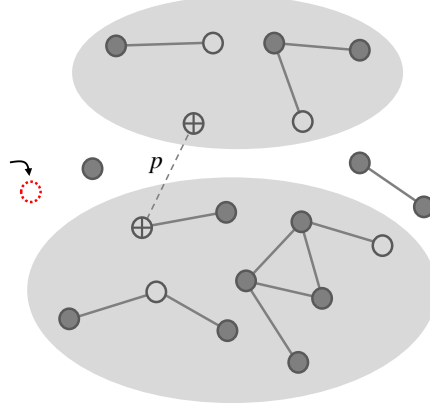


Fig. 2.21: (Color online) Schematic illustration of our model for $m = 3$. Nodes are represented by solid circles. After a new node (red dotted open circle) is added to the system, two sets of m nodes (solid open circles) are randomly selected from distinct clusters. For each set represented by a shaded ellipse, the node belonging to the smallest cluster (represented by \oplus) is chosen. The two nodes \oplus are connected with probability p by a link (dashed line).

by applying the local suppression rule, where two nodes belonging to the two smallest clusters among m randomly selected nodes are connected with probability p . In the unified framework, we derive the analytic solutions of all these models for growing and static networks.

Adopting the notation of the da Costa model [19, 39], we define $P_m(s, p, t)$ as the probability that a selected node belongs to the cluster of size s at time t for a given control parameter m representing the strength of suppression, where p denotes the probability that a link is added between the two selected nodes. Then the rate equation of $P_m(s, p, t)$ is written as

$$\frac{d}{dt} \left(N(t) P_m(s, p, t) \right) = sp \left[\sum_{u+v=s} Q_m(u, p, t) Q_m(v, p, t) - 2Q_m(s, p, t) \right] + \delta_{1s}, \quad (2.5)$$

where

$$Q_m(s, p, t) = \sum_{k=1}^m \binom{m}{k} P_m(s, p, t)^k \left[1 - \sum_{u=1}^s P_m(u, p, t) \right]^{m-k}. \quad (2.6)$$

$Q_m(s, p, t)$ is the probability that the smallest cluster among the m clusters to which the m randomly chosen nodes belong at time t is of size s for a given p . The last term, δ_{1s} , in Eq. (2.5) indicates that a new node of size one is added to the system at each time step. For static networks, the last term disappears, and the total number of nodes $N(t)$ is fixed at constant N . Moreover, the linking probability p is unity because a link is always added at each time step in the static network model. The above rate equation of $P_m(s, p, t)$ is equivalent to that in Refs. [19, 39] with the time normalized by the system size N .

2.2.2 Scaling relations of critical exponents

Here, we try to determine the scaling relations using the scaling forms of $P_m(s, p, t)$ and $Q_m(s, p, t)$ in growing networks. In the steady state limit, i.e., $N \gg 1$ and $t \gg 1$, assuming that $P_m(s, p, t)$ and $Q_m(s, p, t)$ are independent of time t , they thus can be written as $P_m(s, p)$ and $Q_m(s, p)$. Then Eq. (2.5) becomes

$$P_m(s, p) = sp \left(\sum_{u+v=s} Q_m(u, p) Q_m(v, p) - 2Q_m(s, p) \right) + \delta_{1s}. \quad (2.7)$$

As p is increased, cluster formation becomes more likely. Numerical simulations [26] show that, above the percolation threshold p_c , a percolat-

ing cluster of size G emerges as $G \sim (p - p_c)^\beta$ for $m \geq 2$. The two distributions, $P_m(s, p)$ and $Q_m(s, p)$, satisfy the sum rules $\sum_s P_m(s, p) = 1 - G$ and $\sum_s Q_m(s, p) = 1 - G^m$, where an infinite cluster is excluded from the sums. The n -th moments of the cluster sizes for each distribution are expressed as $\langle s^n \rangle_P = \sum_s s^n P_m(s, p)$ and $\langle s^n \rangle_Q = \sum_s s^n Q_m(s, p)$. Eq. (2.7) for finite components leads to the following equations:

$$G = 2pG^m \langle s \rangle_Q, \quad (2.8)$$

$$\langle s \rangle_P = 2p \langle s \rangle_Q^2 - 2pG^m \langle s^2 \rangle_Q + 1. \quad (2.9)$$

Next, $P_m(s, p)$ is assumed to follow scaling behavior near p_c as

$$P_m(s, p) = s^{1-\tau} f(s/s_c), \quad (2.10)$$

where s_c is a characteristic cluster size and behaves as $s_c \sim |p - p_c|^{-1/\sigma}$. In addition, $f(x)$ is a scaling function that by definition is constant for $x \ll 1$ and decays exponentially for $x \gg 1$. From this, we obtain that $\beta = (\tau - 2)/\sigma$.

Replacing the summation in Eq. (2.6) with an integral, we find

$$Q_m(s, p) \cong m \left(\int_s^\infty du P_m(u, p) \right)^{m-1} P_m(s, p) \quad (2.11)$$

for large s in the steady state limit. Then the scaling form of $Q_m(s, p)$ is

obtained as follows:

$$Q_m(s, p) = s^{(2m-1)-m\tau} g(s/s_c), \quad (2.12)$$

where $g(x)$ is a scaling function of $Q_m(s, p)$, corresponding to $f(x)$ for $P_m(s, p)$.

Because the first moments of the cluster sizes diverge at the critical point as $\langle s \rangle_P \sim (p - p_c)^{-\gamma_P}$ and $\langle s \rangle_Q \sim (p - p_c)^{-\gamma_Q}$, Eqs. (2.10) and (2.12) produce the following two scaling relations:

$$\gamma_P = (3 - \tau)/\sigma, \quad (2.13)$$

$$\gamma_Q = (2m + 1 - m\tau)/\sigma. \quad (2.14)$$

Moreover, plugging $\langle s \rangle_P$ and $\langle s \rangle_Q$ into Eqs. (2.8) and (2.9), we obtain that

$$\gamma_P = 2\gamma_Q = 2(m - 1)\beta. \quad (2.15)$$

By using Eqs. (2.13)–(2.15), the explicit forms of the critical exponents γ_P , γ_Q , $1/\sigma$, and τ are obtained in terms of β and m as follows:

$$\gamma_P = 2(m - 1)\beta, \quad (2.16)$$

$$\gamma_Q = (m - 1)\beta, \quad (2.17)$$

$$\frac{1}{\sigma} = (2m - 1)\beta, \quad (2.18)$$

$$\tau = 2 + \frac{1}{2m - 1}. \quad (2.19)$$

We remark that these formulas differ from the corresponding formulas for static network [39]. The two exponent formulas for the static and growing cases are compared in Table 2.5. We also note that the four formulas above are consistent with those obtained in the previous study [26] of the minimal rule, but m is replaced by $2m$, because the minimal rule chose m nodes randomly, and not $2m$ nodes as in this model. Finally, we remark that the exponent τ is independent of β for the growing model but depends on β for the static model.

In the supercritical regime, $p > p_c$, where the giant cluster emerges, Eq. (2.11) can be simply approximated as $Q_m(s, p) \cong mG^{m-1}P_m(s, p)$. The generating functions of $P_m(s, p)$ and $Q_m(s, p)$ are introduced as

$$\mathcal{P}_m(z, p) \equiv \sum_{s=1}^{\infty} P_m(s, p) z^s$$

and

$$\mathcal{Q}_m(z, p) = \sum_{s=1}^{\infty} Q_m(s, p) z^s,$$

respectively. The relation between the two generating functions can be written as

$$\begin{aligned} 1 - G^m - \mathcal{Q}_m(z, p) &= \sum_s Q_m(s, p) [1 - z^s] \\ &\cong \sum_s mG^{m-1} P_m(s, p) [1 - z^s] \\ &= mG^{m-1} [1 - G - \mathcal{P}_m(z, p)]. \end{aligned} \tag{2.20}$$

Therefore,

$$1 - \mathcal{Q}_m(z, p) = mG^{m-1} \left[1 - \mathcal{P}_m(z, p) - \frac{m-1}{m}G \right], \quad (2.21)$$

where the sum rules $\sum_s \mathcal{Q}_m(s, p) = 1 - G^m$ and $\sum_s \mathcal{P}_m(s, p) = 1 - G$ are applied. Then, Eq. (2.7) becomes

$$\mathcal{P}_m(z, p) = 2m^2 G^{2(m-1)} p \times \left[\mathcal{P}_m(z, p) - 1 + \frac{m-1}{m}G \right] \frac{\partial \mathcal{P}_m(z, p)}{\partial \ln z} + z. \quad (2.22)$$

When $z = 1$, the Eq. is

$$\mathcal{P}_m(1, p) - 1 = 2m^2 G^{2(m-1)} p \times \left[\mathcal{P}_m(1, p) - 1 + \frac{m-1}{m}G \right] \langle s \rangle_P. \quad (2.23)$$

Using the relations $G(p) \sim (p - p_c)^\beta$ and $\langle s \rangle_P \sim (p - p_c)^{-\gamma_P}$, one obtains $\gamma_P = 2(m-1)\beta$ again, which is consistent with Eq. (2.16).

2.2.3 Analytic solution of the transition point

To determine the transition point p_c , we derive the scaling functions of $f(x)$ and $g(x)$ with respect to x . First, by substituting $\mathcal{Q}_m(u, p) = \mathcal{Q}_m(s, p) + [\mathcal{Q}_m(u, p) - \mathcal{Q}_m(s, p)]$ into Eq. (2.7), one obtains

$$\begin{aligned} P_m(s, p) = p & \left[-s(s-1)Q^2(s) + 2sQ_m(s) \left(1 - \sum_{u=s}^{\infty} Q_m(u) \right) \right. \\ & \left. + s \sum_{u=1}^{s-1} (Q_m(u) - Q_m(s)) (Q_m(s-u) - Q_m(s)) - 2sQ_m(s) \right] + \delta_{1s}. \end{aligned} \quad (2.24)$$

In the integral form for large s , this equation becomes

$$P_m(s, p) \cong p \left[-s^2 Q^2(s) - 2s Q_m(s) \int_s^\infty Q_m(u) du \right. \\ \left. + s \int_0^s (Q_m(u) - Q_m(s)) (Q_m(s-u) - Q_m(s)) du \right]. \quad (2.25)$$

The scaling form of $P_m(s, p)$ for large s in the critical region is $P_m(s, p) = s^{1-\tau} f(s\delta^{1/\sigma}) = \delta^{(\tau-1)/\sigma} \tilde{f}(s\delta^{1/\sigma})$, where $\delta = |p - p_c| \ll 1$. The scaling form of $Q_m(s, p)$ is $Q_m(s, p) = s^{(2m-1)-m\tau} g(s\delta^{1/\sigma}) = \delta^{[m\tau-(2m-1)]/\sigma} \tilde{g}(s\delta^{1/\sigma})$. We obtain the following equation for the scaling functions.

$$\tilde{f}(x) = p_c \left[-x^2 \tilde{g}^2(x) - 2x \tilde{g}(x) \int_x^\infty dy \tilde{g}(y) \right. \\ \left. + x \int_0^x dy [\tilde{g}(y) - \tilde{g}(x)] [\tilde{g}(x-y) - \tilde{g}(x)] \right], \quad (2.26)$$

where $x \equiv s\delta^{1/\sigma}$, and $(2m-1)(\tau-2) = 1$. This relation is also consistent with Eq. (2.19). Using Eqs. (2.10) and (2.12), we can obtain the following equation:

$$\tilde{g}(x) = m \left[\int_x^\infty dy \tilde{f}(y) \right]^{m-1} \tilde{f}(x), \quad (2.27)$$

where $g(x) = x^{m\tau-(2m-1)} \tilde{g}(x)$, and $f(x) = x^{\tau-1} \tilde{f}(x)$. These relations are all valid for the normal phase, $p < p_c$. For the percolating phase, $p > p_c$, Eqs. (2.26) and (2.27) are valid after the signs of each term that contains $\tilde{f}(x)$ are reversed.

Now, we assume that $f(x)$ and $g(x)$ are expandable for small x around

0 as follows:

$$f(x) = f(0) + a_1 x^\sigma + a_2 x^{2\sigma} + \dots, \quad (2.28)$$

$$g(x) = g(0) + b_1 x^\sigma + b_2 x^{2\sigma} + \dots. \quad (2.29)$$

When Eqs. (2.28) and (2.29) are substituted into Eqs. (2.26) and (2.27), the relation between $f(0)$ and $g(0)$ becomes

$$f(0) + O(x^\sigma) = p_c \left(\frac{g^2(0) \Gamma[-m(\tau-2)]^2}{\Gamma[-2m(\tau-2)]} + O'(x^\sigma) \right), \quad (2.30)$$

$$g(0) = \frac{m}{(\tau-2)^{m-1}} f^m(0), \quad (2.31)$$

where $O(x^\sigma)$ and $O'(x^\sigma)$ represent the higher-order terms of x^σ . Unlike the Eq. for static networks, Eq. (2.7) contains the factor p explicitly. Thus, the transition point p_c can be determined by comparing the zeroth-order term of Eq. (2.30) together with Eq. (2.31) as follows:

$$p_c = \frac{f(0)^{1-2m} \Gamma[-2m(\tau-2)]}{m^2 (\tau-2)^{2-2m} \Gamma[-m(\tau-2)]^2}, \quad (2.32)$$

where $\Gamma(z)$ is a gamma function defined as $\Gamma(z) \equiv (z-1)!$. When Eq. (2.19) is substituted into Eq. (2.32), the dependence of the critical exponent τ disappears and p_c is obtained as follows:

$$\begin{aligned} p_c &= \frac{f(0)^{1-2m} \Gamma[-2m/(2m-1)]}{m^2 (2m-1)^{2m-2} \Gamma[-m/(2m-1)]^2} \\ &= \frac{f(0)^{1-2m}}{2m(2m-1)^{2m-2}} B^{-1} \left[\frac{m-1}{2m-1}, \frac{m-1}{2m-1} \right], \end{aligned} \quad (2.33)$$

where $B^{-1}(x, y)$ is the inverse of the beta function and follows the relationship $B^{-1}(x, y) = \Gamma(x+y)/[\Gamma(x)\Gamma(y)]$. The transition point thus depends only on $f(0)$ and m . As shown in Ref. [38], the value of $f(0)$ can be estimated. Assuming that $P_m(s, p_c)$ follows a power-law function such as $P_m(s, p_c) \sim f(0)s^{1-\tau}$ for all cluster sizes s larger than another characteristic size s_0 , we can write the normalization by $P_m(s, p)$ as follows.

$$\sum_{s=1}^{\infty} P_m(s, p_c) = \sum_{s < s_0} P_m(s, p_c) + f(0) \sum_{s \geq s_0} s^{1-\tau} = 1. \quad (2.34)$$

To solve Eq. (2.34) for $m = 2$, we plot $f(0)$ versus $1/s_0$ in Fig. 2.22 and estimate $f(0)$ to be $\approx 0.217(1)$; $p_c^* = 0.515(1)$. For various m values between 2 and 10, p_c^* is obtained using Eq. (2.32); the results are listed with

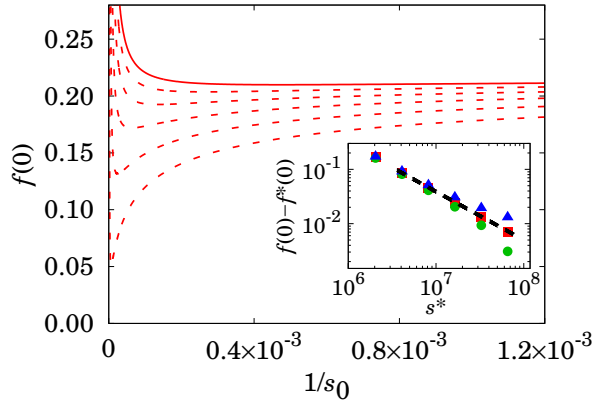


Fig. 2.22: (Color online) Plot of $f(0)$ versus $1/s_0$ for $m = 2$ with s_0 for truncated cluster sizes $s^*/10^3 = 2^{11}, 2^{12}, 2^{13}, 2^{14}$, and 2^{15} (red dashed curves from bottom to top), and 2^{16} (red solid curve). Inset: The corresponding plot of $f(0) - f^*(0)$ versus s^* at $p = 0.51515$ around the transition point p_c , where $f^*(0)$ is the minimum value of $f(0)$ for a given s^* . Data points are for trial values of $f^*(0) = 0.213$ (●), 0.217 (■), and 0.223 (▲). The black dashed line is a guide line and $f(0)$ is estimated as $0.217(1)$ for $m = 2$.

the corresponding $f(0)$ values in Table 2.3. They are consistent with those obtained from the rate equations within the error bars. Moreover, we find that $f(0)$ decays asymptotically as $1/(2m + 0.15)$.

Now, to investigate the asymptotic behavior of the percolation transition point p_c , we consider the Taylor expansion of Eq. (2.33) around $1/m = 0$. For $m \gg 1$, we can use the approximations $B^{-1} \left[\frac{m-1}{2m-1}, \frac{m-1}{2m-1} \right] \approx 1/\pi - (\gamma + \psi(\frac{1}{2}))/2\pi m$ and $(2m + 0.15)^{2m-1} / [2m(2m-1)^{2m-2}] \approx (3.14 - \frac{2.62}{m})$, where γ is the Euler-Mascheroni constant, and $\psi(z)$ is the zeroth-order polygamma function following the relation $\psi(z) = \Gamma'(z)/\Gamma(z)$. Substituting these approximations into Eq. (2.33), we derive the asymptotic behavior of $1 - p_c \sim 1/m$, which decreases algebraically as m is increased.

2.2.4 Numerical solutions of the rate equation

Here, we check numerically the analytic result for the transition point p_c and the scaling relations, and obtain various critical exponent values. To this end, we first obtain $P_m(s, p)$ from the rate equation in Eq. (2.7) up to the order of s^* explicitly [48]. Here, s^* is taken as large as possible for numerical accuracy, but it should be less than s_c . Then, we determine the value of p_c as that at which $P_m(s, p_c)$ exhibits power-law decay with respect to s [1, 2, 26, 40]. This criterion is valid for a second-order percolation transition. Second, we determine the exponent τ by measuring the slope of $\ln P_m(s, p_c)$ with respect to $\ln s$, because the slope is $1 - \tau$. Third, we determine the exponent $1/\sigma$ by plotting $P_m(s, p_c)s^{\tau-1}$ versus $s|p - p_c|^{1/\sigma}$ for different p values. With an appropriate choice of σ , plots for different p values can be collapsed onto a single curve. Next, to determine the exponent

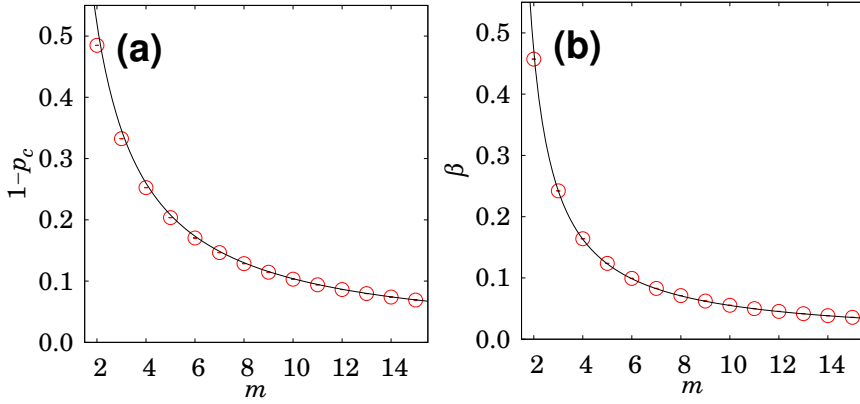


Fig. 2.23: (Color online) (a) Plot of $1 - p_c$ versus m . Data points obtained from the rate equation seem to be fitted by the formula $1 - p_c = 1.04/m$. (b) Plot of β versus m . Data points obtained from the rate equation are fitted by the formula $\beta = 1/(2m - 1.87)$. The critical exponent β decreases algebraically with increasing m .

β , we plot $G(p)$ using $1 - \sum_{s=1}^{s^*} P_m(s, p)$ versus $p - p_c$ ($p > p_c$) on the double logarithmic scale. We then measure the slope as β . Similarly, we obtain the values of the exponents γ_p and γ_Q .

The estimated transition points and critical exponents for $m = 2, \dots, 10$ are presented in Table 2.3. We find that the transition point p_c and exponents β seem to behave as $1 - p_c = 1.04/m$ and $\beta = 1/(2m - 1.87)$, respectively, as shown in Fig. 2.23. Moreover, the estimated values of the critical exponents τ , $1/\sigma$, γ_p , and γ_Q seem to satisfy the scaling relations in Eqs. (2.16)–(2.19), as shown in Fig. 2.24.

We also check the critical exponents and transition point p_c by direct simulations. We grow the networks to $N(t) = 2^{10} \times 10^4$ and repeat this growth more than 10^4 times for $m = 2$, as shown in Fig. 2.25. Using the finite-size scaling approach, $G_N(p) = N^{-\beta/\bar{\nu}} f((p - p_c)N^{1/\bar{\nu}})$ [1, 2, 26], we

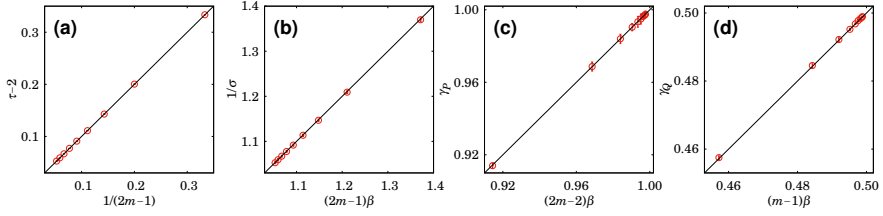


Fig. 2.24: (Color online) Test of the formulas for the exponents (a) $\tau = 2 + 1/(2m - 1)$, (b) $1/\sigma = (2m - 1)\beta$, (c) $\gamma_P = 2(m - 1)\beta$, and (d) $\gamma_Q = (m - 1)\beta$ using numerical data obtained from the rate equation. The error bars are presented in (a), (b), and (d). Data are fitted to the straight solid line, $y = x$, following Eqs. (2.16)–(2.19) in our growing network model.

find $\beta/\bar{v} = 0.175 \pm 0.010$, and $1/\bar{v} = 0.383 \pm 0.010$; thus $\beta = 0.458 \pm 0.038$ for $m = 2$. Using $v = 1/2$ in the mean-field limit, the hyperscaling relation becomes $d_u - 2 = 4\beta^*$ for a given m [19, 39], where d_u represents the upper critical dimension and β^* is the critical exponent of the so-called observable order parameter. In our model, the observable order parameter is G^m in the thermodynamic limit as $t \rightarrow \infty$, because the probability that a node chosen under an aggregation rule belongs to a giant cluster acts as an observable or-

Table. 2.3: Numerical estimates of the percolation threshold p_c , the exponent of the order parameter β , the exponent of the cluster size distribution τ , the exponent of the characteristic cluster size σ , and the exponents of the susceptibility γ_P , γ_Q of growing network models for $m = 2$ to 10. The transition point p_c obtained from the rate equations is compared with p_c^* , which was analytically solved using the scaling functions. $f(0)$ is the coefficient of the leading term of $P_m(s, p) \approx f(0)s^{-\tau}$ for large s . We confirm that p_c and p_c^* are consistent with each other within errors.

m	$f(0)$	p_c^*	p_c	β	τ	σ	γ_P	γ_Q
2	0.217(1)	0.515(1)	0.515(1)	0.457(1)	2.333(1)	0.730(2)	0.914(2)	0.458(1)
3	0.157(1)	0.666(1)	0.667(1)	0.242(1)	2.200(1)	0.827(2)	0.969(3)	0.484(1)
4	0.120(1)	0.745(1)	0.747(1)	0.164(1)	2.143(1)	0.872(2)	0.984(3)	0.492(1)
5	0.097(1)	0.795(1)	0.796(1)	0.124(1)	2.111(1)	0.898(2)	0.990(2)	0.495(1)
6	0.082(1)	0.830(1)	0.830(1)	0.099(1)	2.091(1)	0.916(2)	0.993(3)	0.497(1)
7	0.070(1)	0.854(1)	0.853(1)	0.083(1)	2.077(1)	0.928(2)	0.995(3)	0.498(1)
8	0.062(1)	0.868(2)	0.871(1)	0.071(1)	2.067(1)	0.937(2)	0.996(2)	0.498(1)
9	0.055(1)	0.881(3)	0.885(1)	0.062(1)	2.059(1)	0.944(2)	0.997(2)	0.499(1)
10	0.050(1)	0.900(2)	0.897(1)	0.055(1)	2.053(1)	0.950(2)	0.998(2)	0.499(1)

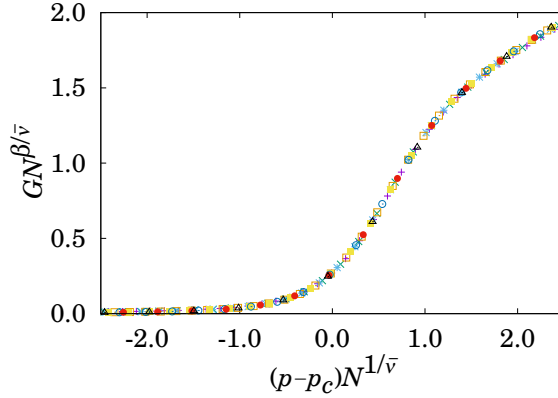


Fig. 2.25: (Color online) Finite-size scaling of $GN^{\beta/\bar{v}}$ versus $(p - p_c)N^{1/\bar{v}}$ from Monte Carlo simulations with more than 10^4 realizations for system sizes $N/10^4 = 2^3, \dots, 2^{10}$ for $m = 2$. Data collapse onto a single curve with $p_c = 0.515(1)$, $1/\bar{v} = 0.383(10)$, and $\beta/\bar{v} = 0.175(10)$. One then get $\beta = 0.458(38)$.

der parameter. Our rule selects the node that belongs to the smallest of the m candidate clusters; the probability that this node is in the giant cluster is G^m . Thus, the observable order parameter exponent is $\beta^* = m\beta$ in our model and the hyperscaling relation ultimately becomes $d_u - 2 = 4m\beta$ for general values of m . For $m = 2$, we obtain $d_u = 5.66 \pm 0.30$ and the correlation volume exponent $d_u v = 2.83 \pm 0.15$, which are consistent with the value obtained

Table. 2.4: Values of the upper critical dimension d_u obtained from the hyperscaling relation in growing networks for $m = 2 - 5$. The correlation volume exponent $d_u v$ for the mean-field theory values $v = 1/2$ and numerically estimated \bar{v} from the finite-size scaling approach is consistent with simulation data within errors. The network is grown to $N = 2^{10} \times 10^4$, and the ensemble average is taken over more than 10^4 samples for each m .

m	d_u	$d_u v$	\bar{v}	β	$1/\bar{v}$
2	5.66(30)	2.83(15)	2.61(7)	0.458(38)	0.383(10)
3	4.91(36)	2.46(18)	2.42(6)	0.243(30)	0.413(10)
4	4.63(42)	2.32(21)	2.28(5)	0.165(27)	0.439(10)
5	4.49(50)	2.24(25)	2.21(5)	0.124(25)	0.452(10)

by simulations and the finite-size scaling approach, $d_u v = \bar{v} = 2.61 \pm 0.07$, within errors. Therefore, the hyperscaling relation holds, and the upper critical dimensions in the growing network depend on m , but their numerical values differ from those of the static network model. Similarly, the hyperscaling relations are tested for different m between 2 and 5, and the corresponding values of d_u are presented in Table 2.4. Finally, the analytical formula of d_u is summarized in Table 2.5.

2.2.5 Analysis

With regard to growing networks, we confirmed that the local suppression effect changed the type of percolation transition from infinite order to second order. Subsequently, we analytically derived the critical exponents τ and $1/\sigma$ for the probability of selecting a node in a cluster of size s , $P_m(s, p) \sim s^{1-\tau} f(s/s^*)$, where $s^* \sim (p - p_c)^{-1/\sigma}$ in terms of a control parameter m , representing the suppressing strength. Furthermore, transition point p_c and other critical exponents were obtained in terms of m . They are summarized in Table 2.5 and compared with those in static networks [19, 26, 39]. Our findings were confirmed by numerically solving the rate equations.

Interestingly, we discovered that as $m \rightarrow \infty$, the transition point and critical exponents behaved as $p_c \rightarrow 1$, $\tau \rightarrow 2$, $\sigma \rightarrow 1$, $\beta \rightarrow 0$, $\gamma \rightarrow 1$, and the upper critical dimension $d_u \rightarrow 4$. The fact that $\beta = 0$ as $m \rightarrow \infty$ indicated that the percolation transition was discontinuous because the suppression effect became global [20, 21]. We remark that, for growing networks, p_c and the critical exponents algebraically approached their respective asymptotes as

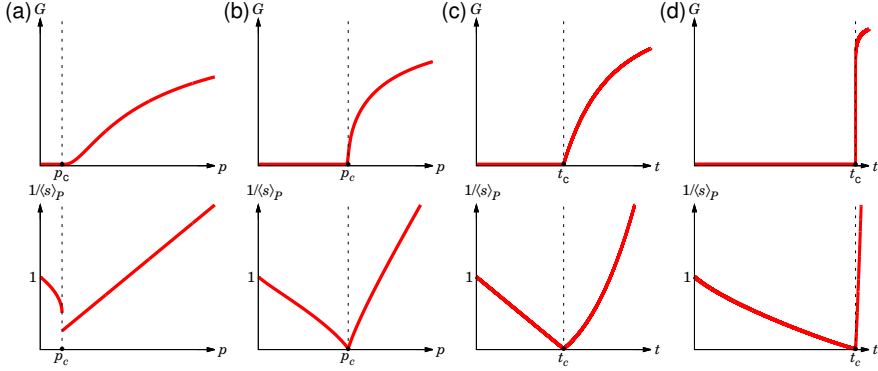


Fig. 2.26: (Color online) Schematic plots of the order parameter G and the inverse first moment of $P_m(s, p)$, $1/\langle s \rangle_P$, for the (a) GRN, (b) d -GRN/ m -GRN, (c) ER, and (d) d -ER/ m -ER models. Schematic plots for the m -GRN and m -ER models are very similar to those for the d -GRN and d -ER models. The only difference between the two suppression rules is that twice as many nodes are selected under the da Costa rule as under the minimal rule. Thus, $m = 2$ under the minimal rule and $m = 1$ under the da Costa rule are reduced to the GRN model in growing networks and the ER model in static networks, respectively.

$m \rightarrow \infty$, whereas for static networks, they exponentially approached them as $m \rightarrow \infty$ and $d_u \rightarrow 2$. Accordingly, for a specified finite m , the order parameter G increases slowly in growing networks, whereas it increases drastically in static networks, as shown in Fig. 2.26.

The results we obtained in this study have led us to reinterpret the original results [16] regarding explosive percolation transitions from a new perspective. In the original paper, the AP was applied to static random networks under the product rule, where an edge minimizing the product of the sizes of merged components is selected between two selected random edges. The order parameter increased drastically even when only two candidate edges were used, which may correspond to $m = 2$ in our AP rules. Hence, the explosive percolation transition type was regarded as a discontinuous tran-

Table. 2.5: Transition point $p_c(t_c)$; the order-parameter exponent β ; the scaling relations of the critical exponents τ , $1/\sigma$, and γ_p ; and the upper critical dimension d_u in growing (static) networks under two different suppression rules for candidate node selection. The minimal rule for the growing and static networks and the da Costa rule in the static networks are compared. The difference between these two suppression rules is that twice as many nodes are selected under the da Costa rule as under the minimal rule.

Network	Process	p_c, t_c	β	τ	$1/\sigma$	γ_p	d_u
Growing network	minimal rule [26]	$1 - \frac{1.81}{m}$	$\frac{1}{m-1.56}$	$2 + \frac{1}{m-1}$	$(m-1)\beta$	$(m-2)\beta$	$2 + 2m\beta$
	da Costa rule	$1 - \frac{1.04}{m}$	$\frac{1}{2m-1.87}$	$2 + \frac{1}{2m-1}$	$(2m-1)\beta$	$2(m-1)\beta$	$2 + 4m\beta$
Static network	minimal rule [26]	$1 - 0.4e^{-0.59m}$	$0.5e^{-0.70m}$	$2 + \frac{\beta}{1+(m-1)\beta}$	$1 + (m-1)\beta$	$1 + (m-2)\beta$	$2 + 2m\beta$
	da Costa rule [19, 39]	–	$1.0e^{-1.43m}$	$2 + \frac{\beta}{1+(2m-1)\beta}$	$1 + (2m-1)\beta$	$1 + 2(m-1)\beta$	$2 + 4m\beta$

sition in the early stages. In retrospect, this hasty conclusion might have been made because the critical exponent β of the order parameter decayed exponentially to zero with increasing m for static networks, even though β was still finite for a finite m . If the explosive percolation model had been considered with random growing networks, then such a conclusion would not have been made.

Chapter 3

Global suppression effects in growing networks

3.1 Model: r -GRN model

The r -GRN model starts with a single node. At each time step, a new node is added to the system. Thus, the total number of nodes at time step t becomes $N(t) = t + 1$. As time goes on, clusters of connected nodes form. At each time step, we classify clusters into two sets, set \mathbf{R} and its complement set \mathbf{R}^c , according to their sizes. Set \mathbf{R} contains $\lceil gN(t) \rceil$ nodes belonging to the smallest clusters, whereas set \mathbf{R}^c contains the nodes belonging to the rest large clusters. $g \in [0, 1]$ is a parameter that controls the size of \mathbf{R} . Let c_i denote the i -th cluster in ascending order of cluster size. Suppose that the $(k - 1)$ -th cluster satisfies the condition $\sum_{i=1}^{k-1} s(c_i) < \lceil gN \rceil \leq \sum_{i=1}^k s(c_i)$, where $s(c_i)$ denotes the size of the cluster with index c_i . Then, set $\mathbf{R}(t)$ contains all the nodes belonging to the $k - 1$ smallest clusters and $\lceil gN \rceil - \sum_{i=1}^{k-1} s(c_i)$ nodes randomly selected from the k -th smallest cluster. The complement set \mathbf{R}^c contains the remaining (largest) clusters. Next, one node is selected randomly from set $\mathbf{R}(t)$ and another is selected from among all the nodes. A node in the set of smaller clusters has twice chance of being linked, while a node in the set of larger clusters has one chance. Then, a link

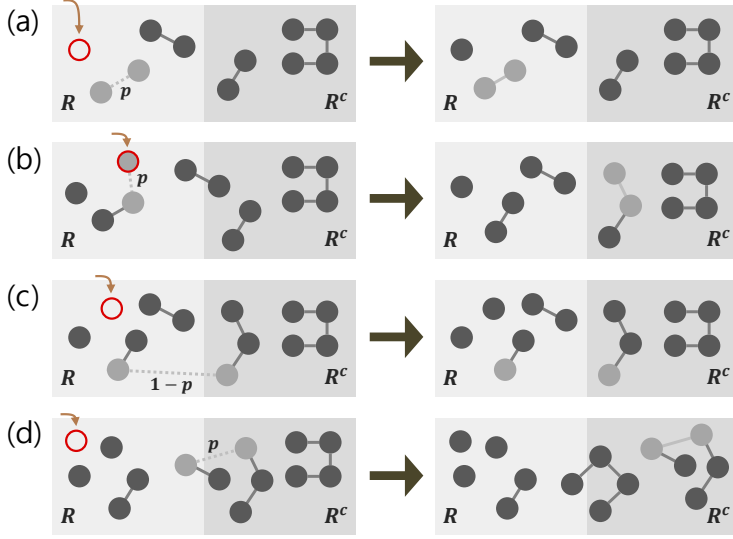


Fig. 3.27: Schematic illustration of the r -GRN model with $g = 0.4$. Nodes are represented by circles. $\mathbf{R}(t)$ is represented by light gray region and $\mathbf{R}^c(t)$ is done by dark grey region. Solid line between two nodes represents a link. In (a), the system starts at five clusters with sizes $(1, 1, 2, 2, 4)$, respectively, and the total number of nodes $N(t) = 10$ with $S_R = 2$ and $\lceil gN \rceil = 4$ at $t = 9$. After one time step, a new node (red open circle) is added, and $N(t)$ becomes 11 with $S_R = 2$ and $\lceil gN \rceil = 5$ at $t = 10$. Two isolated nodes (filled light grey) in \mathbf{R} and are merged. (b) Next, a new node is added, and so $N(t) = 12$, $S_R = 2$, and $\lceil gN \rceil = 5$ at $t = 11$. In this case, one node of the cluster of size two in set \mathbf{R} moves to set \mathbf{R}^c . The newly added node is merged with the cluster of size two in \mathbf{R} . This cluster moves to set \mathbf{R}^c and the cluster of size two on the boundary moves to \mathbf{R} . (c) Next, a new node is added with $N(t) = 13$, $S_R = 2$, and $\lceil gN \rceil = 6$ at $t = 12$. Selected nodes are not connected with probability $1 - p$. (d) Next, a new node is added with $N(t) = 14$, $S_R = 2$, and $\lceil gN \rceil = 6$ at $t = 13$. And just one node of the cluster of size two in set \mathbf{R} moves to set \mathbf{R}^c . A cluster of size two in \mathbf{R} and the cluster of size three in \mathbf{R}^c are merged, belonging to \mathbf{R}^c . The cluster of size four in \mathbf{R}^c lies on the boundary and the cluster of size two on the boundary moves to \mathbf{R} .

is added between the two selected nodes with link occupation probability p . The selection rule becomes global in the process of sorting out the portion of the smallest clusters among all clusters. Moreover, it suppresses the growth of large clusters by allowing less chance to be linked. In Fig. 3.27,

this link connection process is visualized for the restricted fraction $g = 0.4$ as an example. This restriction rule is initially introduced in Ref. [41] and modified in Refs. [42, 49].

We define the size of the largest cluster in set \mathbf{R} as $S_R(p, t)$ for a given p at time t , which determines the size of the boundary cluster(s) between the two sets. It depends on the fraction g [42]. Thus, when $g = 1$, which means that S_R is equal to the size of the giant cluster, denoted as $GN(t)$, this model reduces to the original GRN model [14]. It has been found previously that the GRN model exhibits a continuous infinite-order phase transition at $p_c = 1/8$ [14]. However, when $g \rightarrow 0$, $S_R = 1$, and an isolated node in \mathbf{R} is merged with a node in \mathbf{R}^c with link occupation probability p .

3.2 Cluster size distribution $n_s(p)$

Let us define the cluster number density $n_s(p, t)$ for a given p at time step t as the number of clusters of size s divided by the current number of nodes $N(t)$ at t . In our previous studies [27], we derived the rate equations according to the cluster size s comparing to S_R for the cluster size distribution $N(t)n_s$. For convenient readability, those rate equations are rewritten in the A.1.

Here we solve the rate equation of $n_s(p)$ for a given g . First, when $s = 1$, the rate equation becomes $\dot{n}_1 = -p(1 + \frac{1}{g})n_1 + 1$ for $S_R > 1$ and $\dot{n}_1 = -p(n_1 + 1) + 1$ for $S_R = 1$. Thus, $n_1(p)$ becomes

$$n_1 = \begin{cases} \frac{1}{1+p(1+\frac{1}{g})} & S_R(p) > 1 \quad (p > p_1), \\ \frac{1-p}{1+p} & S_R(p) = 1 \quad (p < p_1). \end{cases} \quad (3.1)$$

The two solutions become the same at $p = (1 - g)/(1 + g)$, as shown in Fig. 3.28(e). This p is denoted as p_1 . For $g = 0.4$, $p_1 = 0.4285714 \dots$

Next, when $s = 2$, the rate equations are as follows: $n_2 = p[(n_1 n_1/g) - 2n_2(1 + 1/g)]$ for $S_R > 2$; $n_2 = p[(n_1 n_1/g) - 2n_2 - (1 - n_1/g)]$ for $S_R = 2$; $n_2 = p(n_1 - 2n_2)$ for $S_R = 1$. We obtain n_2 as follows:

$$n_2 = \begin{cases} \frac{p \frac{n_1^2}{g}}{1+2p(1+\frac{1}{g})} & S_R > 2 \quad (p > p_2), \\ \frac{p[\frac{n_1^2}{g} - (1 - \frac{n_1}{g})]}{1+2p} & S_R = 2 \quad (p_1 < p < p_2), \\ \frac{pn_1}{1+2p} & S_R = 1 \quad (p < p_1). \end{cases} \quad (3.2)$$

Two kinks (crossovers) exist in $n_2(p)$, as shown in Fig. 3.28(f). The position p of the first kink is just p_1 , and that of the second kink is determined by setting n_2 for $S_R > 2$ equal to that for $S_R = 2$. This position is denoted as p_2 . For $g = 0.4$, $p_2 = 0.5653082 \dots$

In general, when $s > 1$, the cluster size distribution $n_s(p)$ can be obtained from the rate equations in the steady state as follows:

$$n_s(p) = \begin{cases} \frac{p \sum_{i,j=1}^{\infty} \frac{in_i jn_j}{g} \delta_{i+j,s} + \delta_{1s}}{1+sp(1+\frac{1}{g})} & s < S_R, \\ \frac{p \left[\sum_{i,j=1}^{\infty} \frac{in_i jn_j}{g} \delta_{i+j,s} - \left(1 - \sum_{k=1}^{S_R-1} \frac{kn_k}{g} \right) \right]}{1+sp} & s = S_R, \\ \frac{p \left[\sum_{j=1}^{\infty} \sum_{i=1}^{S_R-1} \frac{in_i jn_j}{g} \delta_{i+j,s} + \sum_{j=1}^{\infty} \delta_{S_R+j,s} jn_j \left(1 - \sum_{k=1}^{S_R-1} \frac{kn_k}{g} \right) \right]}{1+sp} & s > S_R. \end{cases} \quad (3.3)$$

There exist s kinks on the curve n_s at p_1, \dots, p_s in ascending order of p . The position of the last kink p_s is determined by setting n_s for $S_R > s$ equal

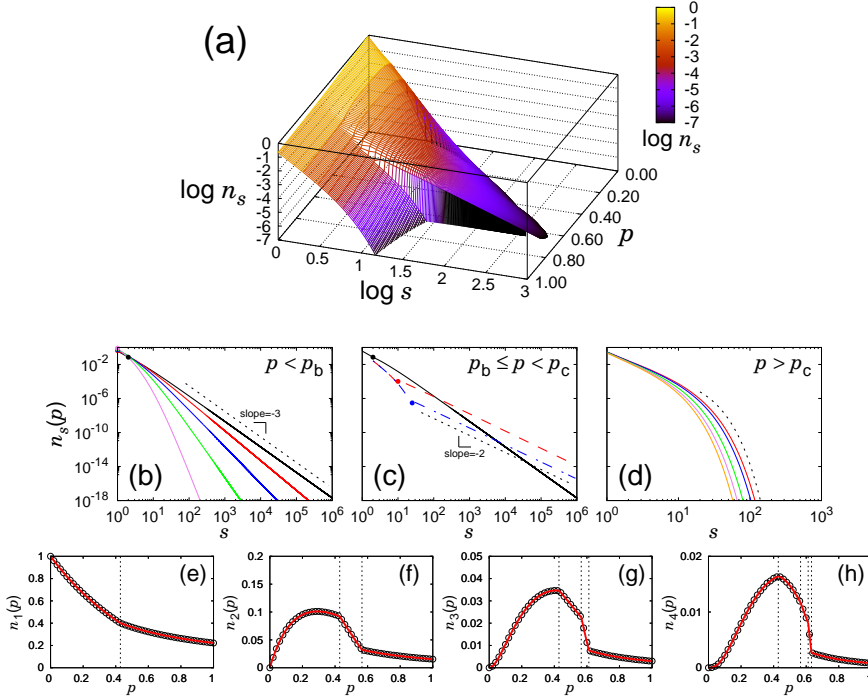


Fig. 3.28: Cluster size distribution $n_s(p)$ as a function of s and p for given g : In this case, $g = 0.4$ are taken. (a) 3D plot of $n_s(p)$ as a function of s and p . Plots (b)-(d) are obtained with several fixed p for $n_s(p)$. (b) For $p < p_b$, $n_s(p)$ follows $\sim s^{-\tau}$ with $\tau > 3$. The slope of the dotted guide line is -3 . Solid lines are obtained for $p = 0.472576 \approx p_b, 0.4, 0.3, 0.2$, and 0.1 from right to left. (c) For $p_b \leq p < p_c$, in the small-cluster-size region, $n_s(p)$ decays exponentially up to S_R and then exhibits power-law decay behavior with $2 < \tau \leq 3$. Solid, dashed, and dashed-dotted lines represent for p_{S_R} with $S_R = 2, 10$ and 25 , respectively. Dotted line is a guide line with slope of -2 . (d) For $p \geq p_c$, $n_s(p)$ exponentially decay. Solid curves represent $n_s(p)$ for $p = 0.6596, 0.7, 0.8, 0.9$, and 1.0 from right to left. Dotted curve is an exponentially decaying guide curve. Plots (e)-(h) are obtained with several values of s for $n_s(p)$. (e) Plot of $n_1(p)$ versus p . A crossover exists at p_1 . (f) Plot of $n_2(p)$ versus p . Two crossover behaviors occur at p_1 and p_2 , where $p_1 < p_2$. (g) and (h) Plots of $n_3(p)$ and $n_4(p)$ versus p , respectively. Symbols represent simulation results, and solid lines are analytical results. Dotted vertical lines represent p_{S_R} for $S_R = 1, 2, 3$, and 4 at $p_{S_R=1} = 0.428571$, $p_{S_R=2} = 0.565302$, $p_{S_R=3} = 0.612016$, and $p_{S_R=4} = 0.632728$.

to n_s for $S_R = s$. For convenience, we use the index as S_R to avoid confusion with the index of cluster size s . The positions p_{S_R} as a function S_R are listed

in Table 3.6. As shown in Figs. 3.28(e)–3.28(h), the interval between two successive crossover points becomes narrower with increasing S_R . The position p_{S_R} seems to converge to a certain value, p_∞ , in a power-law form of $p_\infty - p_{S_R}$ as a function of S_R asymptotically as shown in Fig. 3.29. Here p_∞ is estimated to be 0.65948(1). Figs. 3.28(b)–3.28(d) show the distributions n_s versus s for a given fixed p , which corresponds to the $(\log n_s, \log s)$ plane of the three-dimensional plot of $n_s(p)$ in Fig. 3.28(a).

Table. 3.6: Values of p_{S_R} as a function of S_R for $g = 0.4$.

S_R	p_{S_R}
1	0.4285714285(1)
2	0.5653082407(1)
3	0.6120164684(1)
4	0.6327279058(1)
5	0.6433362667(1)
6	0.6492814220(1)
7	0.6528226406(1)
8	0.6550262003(1)
9	0.6564429142(1)
10	0.6573769871(1)
11	0.6580052394(1)
12	0.6584346536(1)
13	0.6587320681(1)
14	0.6589403439(1)
15	0.6590875632(1)
16	0.6591924579(1)
17	0.6592677124(1)
18	0.6593220275(1)
19	0.6593614370(1)
20	0.6593901656(1)
∞	0.65948(1)

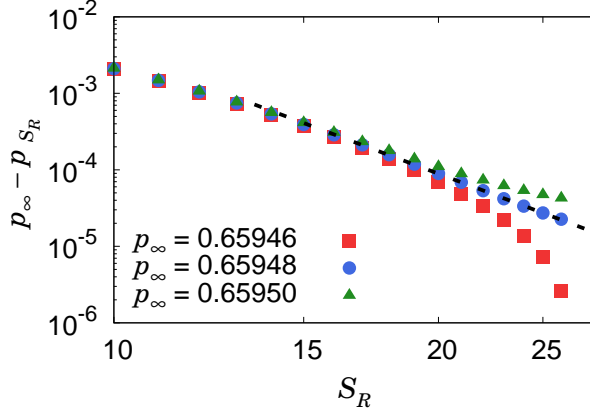


Fig. 3.29: Plot of $p_\infty - p_{S_R}$ versus S_R for $g = 0.4$. When $p_\infty = 0.65948(1)$, a power-law decay appears.

3.3 Two transition points, p_b and p_c

From the cluster size distribution $n_s(p)$ for given p , we find that there exist two transition points, say p_b and p_c , which characterize the following three distinct intervals on the line of p : i) For $p < p_b$, $n_s(p)$ follows the power law $n_s(p) \sim s^{-\tau}$ for $s > S_R$ with exponent $\tau > 3$, whereas it decays exponentially as a function of s for $s < S_R$. ii) For $p_b \leq p < p_c$, $n_s(p)$ also follows a power law with exponent τ for $s > S_R$. Particularly, the exponent τ decreases continuously from $\tau = 3$ to 2 as p is increased from p_b to p_c . For $s < S_R$, $n_s(p)$ decays exponentially as a function of s . iii) For $p > p_c$, a giant cluster is generated and the distribution of the remaining finite clusters decays exponentially as a function of s .

The power-law behavior of $n_s(p)$ with $\tau > 3$ in the region i) is inherited from the infinite-order transition of the GRN model [14]. Thus the region i) is regarded as an infinite-order-type critical region. Meanwhile, in the re-

Table. 3.7: Numerical estimates of the transition points p_b and p_c . The critical exponents τ are calculated at $p = p_b$ and p_c for $g = 0.1 - 0.9$. We note that the exponent τ at p_c becomes difficult to obtain as g approaches one.

g	p_b	p_c	ΔG	$\tau(p_b)$	$\tau(p_c)$
0.1	1/2	0.905(1)	0.900(1)	3.00(1)	2.00(1)
0.2	1/2	0.817(1)	0.800(1)	3.00(1)	2.00(1)
0.3	1/2	0.736(1)	0.700(1)	3.00(1)	2.00(1)
1/3	1/2	0.710(1)	0.666(1)	3.00(1)	2.00(1)
0.4	0.473(1)	0.660(1)	0.600(1)	3.00(1)	2.00(1)
0.5	0.440(1)	0.587(1)	0.500(1)	3.00(1)	2.00(1)
0.6	0.405(1)	0.516(1)	0.400(1)	3.00(1)	2.00(1)
0.7	0.367(1)	0.447(1)	0.300(1)	3.00(1)	1.99(1)
0.8	0.323(1)	0.376(1)	0.200(1)	3.00(1)	1.99(1)
0.9	0.268(1)	0.297(1)	0.100(1)	3.00(1)	1.8(2)
1.0	1/8	1/8	0	3	-

gion ii), because $2 < \tau < 3$, the mean cluster size diverges. Thus the region ii) is regarded as a second-order-type critical region. It is noteworthy that while the critical behavior occurs at a critical point in a prototypical second-order transition, here it occurs in the entire region ii). At p_c^- , $\tau = 2$. This means that clusters are extremely heterogeneous and further suppression of the largest cluster leads to a discontinuous transition. Indeed, a discontinuous transition occurs at p_c . Both transition points for different g values are listed in Table 3.7.

To determine p_b and p_c , here we introduce the generating function $f(x)$ of the probability sn_s that a randomly chosen node belongs to the cluster of size s , defined as

$$f(x) \equiv \sum_{s=1}^{\infty} sn_s x^s, \quad (3.4)$$

where x is the fugacity in the interval $0 < x < 1$. The giant cluster size G is

obtained as $G = 1 - \sum_{s=1}^{\infty} s n_s = 1 - f(1)$. The mean cluster size is obtained as $\langle s \rangle = \sum_{s=1}^{\infty} s^2 n_s = f'(1)$, where the prime represents the derivative with respect to x . To determine p_b (p_c), we consider the case of S_R being finite (infinite).

3.3.1 For finite S_R

When S_R is finite, we derive the recurrence relation for n_s . First, when $S_R = 1$, the rate equations in the steady state are simply reduced as follows:

$$n_1 = -p(n_1 + 1) + 1, \quad (3.5)$$

$$n_s = p[(s-1)n_{s-1} - sn_s] \quad \text{for } s > 1. \quad (3.6)$$

Then, one can obtain the generating function $f(x)$ as

$$f(x) = -xp f'(x) - px + x + px^2 f'(x) + px f(x). \quad (3.7)$$

The giant cluster size G is $G = 1 - \sum_{s=1}^{\infty} s n_s = 1 - f(1) = 0$. The mean cluster size is obtained as $\langle s \rangle = \sum_{s=1}^{\infty} s^2 n_s = f'(1) = 1/(1-2p)$. So the mean cluster size diverges at $p_b = 1/2$. If this value is larger than p_1 for a given g , then we move to $S_R = 2$. When $S_R = 2$, $G = 0$ and $\langle s \rangle = f'(1) = 1/[1-4p+(2pn_1/g)]$.

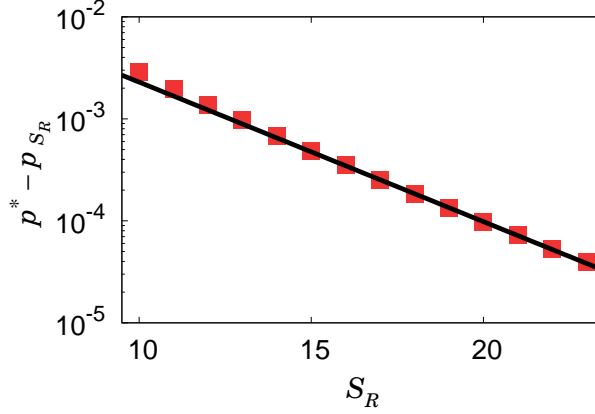


Fig. 3.30: Semi-log plot of $p^* - p_{S_R}$ versus S_R . $p^* - p_{S_R}$ decays exponentially as S_R increases but it's always larger than zero for finite S_R values. So p^* is larger than p_{S_R} for any finite S_R .

Generally, for finite S_R , we obtain the relation

$$\begin{aligned}
 f(x) + xp f'(x) &= x + p \left[\sum_{s=1}^{S_R-1} s x^s \left(-\frac{1}{g} s n_s \right) - S_R x^{S_R} \left(1 - \frac{1}{g} \sum_{s=1}^{S_R-1} s n_s \right) \right. \\
 &\quad + \sum_{s=1}^{S_R-1} \frac{s n_s}{g} x^s (x f'(x) + s f(x)) \\
 &\quad \left. + x^{S_R} (x f'(x) + S_R f(x)) \left(1 - \frac{1}{g} \sum_{s=1}^{S_R-1} s n_s \right) \right]. \quad (3.8)
 \end{aligned}$$

When $x = 1$, Eq. (3.8) may be written as $f(1)J(p) = J(p)$ for the range $p_{S_R-1} \leq p < p_{S_R}$, where $J(p) \equiv 1 - p \left[\sum_{s=1}^{S_R-1} \frac{s^2 n_s}{g} + S_R \left(1 - \sum_{s=1}^{S_R-1} \frac{s n_s}{g} \right) \right]$. Now, let us denote p satisfying $J(p) = 0$ as p^* . We can calculate these values p^* as S_R increases using Eq. (3.3) in the steady state. But p^* is always larger than p_{S_R} so $J(p)$ cannot be zero as shown in Fig. 3.30 for the case $g = 0.4$. Then we can obtain that $f(1) = 1$ for finite S_R and the relative giant cluster $G = 1 - f(1) = 0$.

At $x = 1$, plugging $f(1) = 1$ into the derivative of Eq. (3.8) with respect to x , we obtain that

$$f'(1) = \left[1 + 2p \left(\sum_{s=1}^{S_R-1} \frac{(S_R - s)sn_s}{g} - S_R \right) \right]^{-1} = \langle s \rangle. \quad (3.9)$$

To obtain p_b , once we set $S_R = 1$ and check whether there exists a certain value of p less than p_{S_R} , say p_* , such that $\langle s \rangle^{-1} = 0$. If the solution exists, p_* is a critical point p_b and S_R is the size of the largest cluster in set **R**. Otherwise, we increase S_R by one, and try to find a solution satisfying $\langle s \rangle^{-1} = 0$. We repeat these steps until the solution is found. The obtained values p_b for different g are listed in Table 3.7. The existence of p_b below p_c implies that even though the order parameter $G(p)$ is zero for $p < p_c$, the mean cluster size $\langle s \rangle$ can diverge at p_b before p_c .

3.3.2 For infinite S_R

We consider the limit $S_R(p) = \infty$, which corresponds to the case $p > p_\infty$. In this case, Eq. (A.4) is valid for all cluster sizes s . Eqs. (A.4)–(A.6) reduce to the following two equations:

$$n_1 = \frac{1}{1 + (1 + \frac{1}{g})p}, \quad (3.10)$$

$$n_s = \frac{p}{1 + (1 + \frac{1}{g})sp} \sum_{j=1}^{s-1} \frac{j(s-j)n_j n_{s-j}}{g}, \quad (3.11)$$

where s is limited to finite clusters. The generating function associated with sn_s satisfies the following relation:

$$f(x) = -x\left(1 + \frac{1}{g}\right)pf'(x) + \frac{2}{g}pxf(x)f'(x) + x, \quad (3.12)$$

and in another form,

$$f'(x) = \frac{1 - \frac{f(x)}{x}}{\left(1 + \frac{1}{g}\right) - \frac{2}{g}f(x)} \frac{1}{p}. \quad (3.13)$$

Performing numerical integration, we obtain $f(1)$ and $f'(1)$, which correspond to the order parameter $G(p)$ and $\langle s \rangle$ for given p and g in the region $p \geq p_\infty$.

At p_∞ , this order parameter value $G(p_\infty)$ is not zero but finite, indicating that the transition at p_∞ is first-order. Moreover, $G(p_\infty)$ represents the jump size of the order parameter ΔG of the discontinuous transition. We obtain the cluster size distribution using the Eq. (3.3), which follows a power law with $\tau \simeq 2$. Therefore, we think that $p_\infty = p_c$. The results for G and $1/\langle s \rangle$ in the entire region p are shown in Fig. 3.31 for $g = 0.2, 0.4$, and 0.6 . Numerical data of p_b , p_c , ΔG , $\tau(p_b)$, and $\tau(p_c)$ for different g are listed in Table 3.7. Indeed, the order parameters are discontinuous at p_c for different $g < 1$. We draw a phase diagram shown in Fig. 3.32 in the plane of (p, g) .

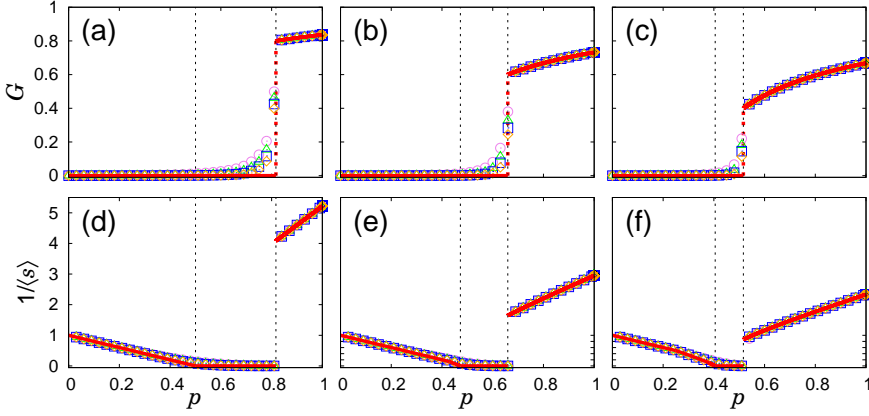


Fig. 3.31: Plot of G and $1/\langle s \rangle$ as a function of p . For $g = 0.2$ in (a) and (d), $g = 0.4$ in (b) and (e), and $g = 0.6$ in (c) and (f), respectively. Symbols represent the simulation results for $N = 10^4$ (\circ), 10^5 (\triangle), 10^6 (\square), and 10^7 (\diamond). Each data point was averaged over 10^3 times. The solid (red) lines are calculated from $f(1)$ and $f'(1)$ for G and $\langle s \rangle$, respectively. The two vertical dotted lines represent p_b and p_c ($p_b < p_c$).

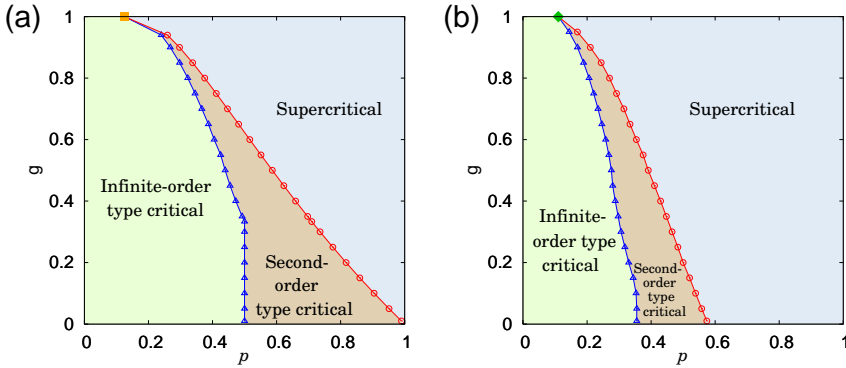


Fig. 3.32: (a) and (b) show the phase diagrams of the r -GRN model and r -PIN model with $\delta = 0.7$, respectively. Symbols \triangle and \circ represent p_b and p_c . $n_s(p)$ decays following a power law with $\tau > 3$ in the infinite-order-type critical region and $2 < \tau < 3$ in the second-order-type critical region. Thus, the mean cluster size is finite and diverges, respectively. As g approaches one, two phase boundaries converge to the conventional transition point $p_c = 1/8$ of the GRN model, represented by \blacksquare , and $p_c = 0.12(1)$ in the PIN model with $\delta = 0.7$, represented by \blacklozenge .

3.4 $\tau(p)$ in the critical region and total number of clusters

When $p < p_c$, the cluster size distribution exhibits a critical behavior: it decays in a power law manner with exponent τ . This exponent τ depends on the link occupation probability p . This property is reminiscent of the feature of the Berezinskii-Kosterlitz-Thouless (BKT) transition in thermal systems. However, the origin of the BKT transition in growing networks differs from that in thermal systems. To illustrate the origin of the critical behavior in growing networks, we consider a limit case with $g \rightarrow 0$ and $S_R = 1$. In this case, cluster merging dynamics occurs only between isolated nodes and another cluster of any size. From Eq. (3.3), one can obtain the explicit form of $n_s(p)$ as follows:

$$n_s(p) = \frac{(s-1)!p^{s-1}n_1(p)}{(1+sp)(1+(s-1)p)\cdots(1+2p)}, \quad (3.14)$$

where $n_1(p)$ is $(1-p)/(1+p)$, and $S_R = 1$. Using the Stirling formula, the gamma function $\Gamma(z) = (z-1)!$ is rewritten as

$$\Gamma(z) \sim z^{z-\frac{1}{2}}e^{-z}\sqrt{2\pi}\left(1 + \frac{1}{12z} + \frac{1}{288z^2} - \frac{139}{51840z^3} - \frac{571}{2488320z^4}\right) \\ \text{as } |z| \rightarrow \infty,$$

one can obtain the asymptotic behavior of Eq. (3.14) as $n_s(p) = \frac{\Gamma(s)\Gamma(\frac{1}{p}+2)}{\Gamma(s+\frac{1}{p}+1)}n_1(p) \sim s^{-(1+\frac{1}{p})}$, where the critical exponent $\tau = 1 + \frac{1}{p}$. Fig. 3.33 shows τ as a function of p . Because the merging dynamics starts from $S_R =$

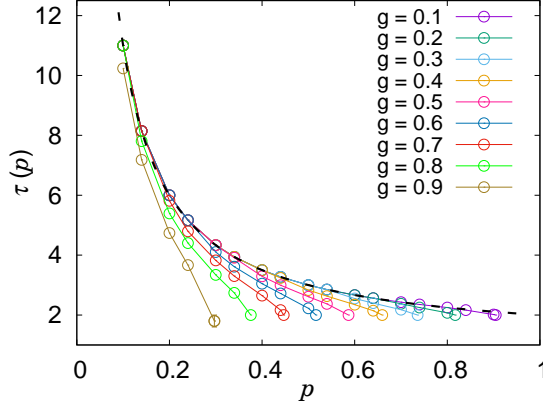


Fig. 3.33: Plot of τ versus p for different g . τ becomes two as p approaches p_c for any g . The black dashed curve is a guide curve representing $1 + 1/p$, which is obtained from the limiting case $S_R = 1$, i.e., $g \rightarrow 0$.

1, $\tau = 1 + 1/p$ appears in the envelope of $\tau(p)$. Thus, addition of a new node into the system at each time step is a key factor that generates the critical region below the transition point p_c .

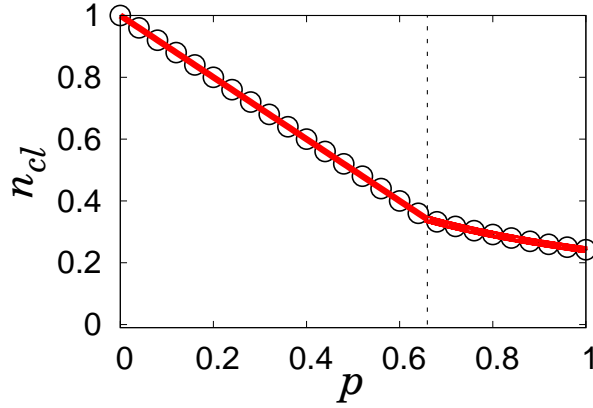


Fig. 3.34: Plot of the total number of clusters n_{cl} versus p for $g = 0.4$. The red solid line is obtained from the rate equation integrating the cluster size distribution. The open circles represent the numerical simulation data for $N = 10^6$, averaged over 10^4 configurations. The black vertical dotted line represents p_c for $g = 0.4$.

The total number of clusters per site, $n_{cl}(p) \equiv \sum_{s=1}^{\infty} n_s(p)$, can be calculated from the rate equations by summing up $n_s(p)$ over all finite clusters in Eq. (3.3). Fig. 3.34 shows $n_s(p)$ for $g = 0.4$. The circle symbols represent $n_s(p)$ obtained from numerical simulations. They are in agreement with theoretical results (solid line) for $g = 0.4$ in the entire p region.

3.5 Universal behavior

Protein interaction network (PIN) models are growing networks and also exhibit the BKT transitions [13]. Nodes in this network represent proteins and links connect functionally related proteins. Connected proteins form a proteome or protein complexes. The proteome network is a usually sparse graph with a small mean degree. Inspired from the biological process, several minimal models for the evolution of PIN were introduced [50]. Here we recall the PIN model proposed in Refs. [13, 51]. The model includes three important features; i) duplication, ii) mutation, and iii) divergence. i) At each time step, a node is newly introduced, which duplicates a randomly chosen nodes (called replicated node) among pre-existing nodes. ii) The node connects to each of the neighbors of the replicated node with probability $1 - \delta$. iii) The new node also can link to all pre-existing node with probability β/N , where N is the current total number of nodes. Thus cluster merging occurs.

In order to apply the global suppression effect of the r -GRN model to the PIN model, we slightly modify the process iii) of the PIN models as follows. Each new node links only to the nodes belonging to set \mathbf{R} with the smallest clusters, which is similarly defined in the r -GRN model. The value

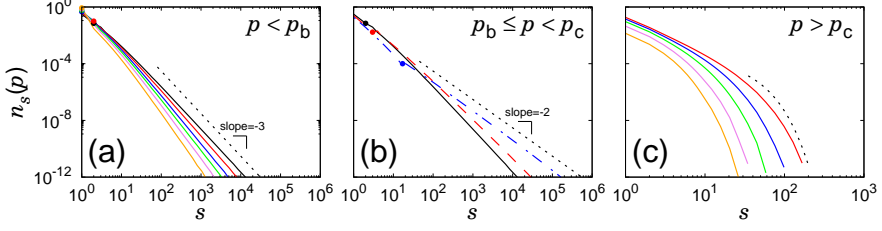


Fig. 3.35: Plots of the cluster size distribution $n_s(p)$ of the r -PIN model as a function of s in different p regions. We binned simulation data logarithmically for $N = 2^{14} \times 10^4$ averaged over 10^3 configurations. $g = 0.4$ and $\delta = 0.7$ are taken. Three cases of $n_s(p)$ are distinguished: (a) For $p < p_b$, $n_s(p)$ asymptotically follows the power law $\sim s^{-\tau}$ with $\tau > 3$. The slope of the dotted guide line is -3 . Solid lines are obtained for $p = 0.29 \approx p_b, 0.25, 0.20, 0.15, 0.10$, and 0.05 from right to left. (b) For $p_b \leq p < p_c$, in the small-cluster-size region, $n_s(p)$ decays exponentially and then exhibits power-law decay behavior with $2 < \tau \leq 3$. Solid (black), dashed (red), and dashed-dotted (blue) lines represent p_{S_R} , where $S_R = 2$ ($p = 0.29$), 3 ($p = 0.35$), and 17 ($p = 0.423$), respectively. Two dotted lines are guidelines with slopes of -2 and -3 . (c) For $p \geq p_c$, $n_s(p)$ for finite clusters shows exponentially decaying distributions. Solid curves represent $p = 0.43, 0.50, 0.60, 0.75$, and 0.90 from right to left. Dotted curve is an exponentially decaying guide curve.

β in the probability β/gN , where gN is the current total number of nodes belonging to set \mathbf{R} , corresponds to p in the r -GRN model. Accordingly, the growth of large clusters is suppressed. From the numerical simulations up to $N = 10^8$ with 1000 ensemble averages, we also observe the abnormal transition behaviors as shown in Fig. 3.35, where the previous BKT transition of the PIN model breaks down but the features of infinite-, second-, and first-order type transitions all occur similarly to the r -GRN model.

For $\delta = 0.7$, we numerically simulate for $g = 0.4$ and $g = 0.6$. Fig. 3.36 shows the two transition points $p_b = 0.29(1)$ and $p_c = 0.43(1)$ in panels (a) and (c) for $g = 0.4$, and $p_b = 0.26(1)$ and $p_c = 0.35(1)$ in panels (b) and (d) for $g = 0.6$. The two transition points obtained from the analytical results are

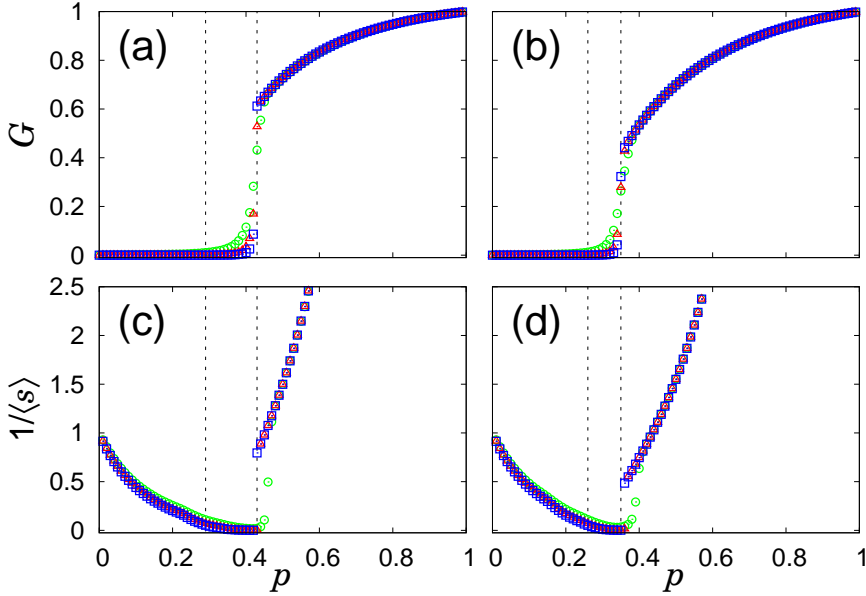


Fig. 3.36: Plot of G and $1/\langle s \rangle$ for the r -PIN model as a function of p for $g = 0.4$ in (a) and (c), and for $g = 0.6$ in (b) and (d), respectively. Symbols represent numerical simulation data for $N = 2^2 \times 10^4$ (\circ), $2^8 \times 10^4$ (\triangle), and $2^{14} \times 10^4$ (\square). Each data point was averaged over 10^2 configurations. The two vertical dotted lines represent p_b and $p_c > p_b$.

consistent with numerical results. This result is also close to that obtained in the r -GRN model. Thus, we argue that our main results are universal independent of detailed model dynamic rules.

3.6 Analysis

When the link occupation probability p is below p_c , most clusters are small and the suppression is not effective. Hence the infinite-order critical behavior of $n_s(p) \sim s^{-\tau(p)}$ appears as the one in the BKT transition. The exponent $\tau(p)$ decreases as p is increased. In the BKT transition, τ decreases down

to three as p is increased to p_c ; however, in this restricted growing random network (r -GRN) model, the exponent $\tau(p)$ can decrease more down to two, because the transition point is delayed by the suppression effect. On the other hand, if the cluster size distribution follows a power law without any exponential cutoff, the largest cluster size scales with the system size $N(t)$ in the steady state as $s_{\max} \sim N^{1/(\tau-1)}$. When τ decreases down to two, the largest cluster grows to the extent of the system size in the steady state. Therefore a discontinuous transition occurs.

As τ decreases below three, the mean cluster size, i.e., the susceptibility is no longer finite. We divide the region $p < p_c$ into two subregions, $p < p_b$ and $p_b < p < p_c$, such that for $p < p_b$, $\tau > 3$, whereas for $p_b < p < p_c$, $2 < \tau < 3$. Thus, the mean cluster size is finite and diverges in the former and latter regions, respectively. Therefore, another type of percolation transition (PT) occurs at p_b . It is interesting to note that the mean cluster size diverges even though the giant cluster does not form yet in the interval $p_b < p < p_c$. That is because the cluster size distribution exhibits a critical behavior without an exponential cutoff. Large clusters still remain in the subextensive size, and they induce heavy fluctuations. We regard the region $p < p_b$ as an infinite-order-type critical region, because it is inherited from the infinite-order transition. The region $p_b < p < p_c$, in which feature of the second-order transition appears, is regarded as the second-order-type critical region. At p_c , a first-order PT occurs. For $p > p_c$, the size distribution of finite clusters does not follow a power law. The region $p \geq p_c$ is regarded as a supercritical region. Therefore, when the infinite-order BKT transition is broken by the suppression effect, a first-order PT occurs; a second-order

critical phase appears; and an infinite-order critical phase still remains. We remark that to the best of our knowledge, this is the first observation of the first-order PT in random growing networks.

The r -GRN model was built based on the restricted Erdős-Rényi (r -ER) model recently introduced in Refs. [42, 49]. This r -ER model is a static network model, containing N nodes all the times. The two-node selection rule for a link connection is the same as that of the r -GRN model but once the two nodes are selected at time step t , they are connected definitely. This model contains a global suppression dynamics. In this r -ER model, a power-law behavior of $n_s(t_c)$ without any exponential cutoff appears only at the point t_c^+ just after the order parameter jumps. The exponent τ is in the range $2 < \tau \leq 5/2$ depending on the parameter g . Thus, the model exhibits not only a discontinuous transition but also a critical behavior. The critical behavior appears in the region where the order parameter is finite [42]. Contrary to the transition behavior of this r -ER static network model, the critical behavior in the r -GRN model appears below the transition point p_c , so that the order parameter still remains zero. These behaviors are depicted schematically in Fig. 3.37.

For the r -GRN model, the power-law decay of $n_s(p)$ appears in a steady state over all cluster sizes without forming any bump or exponential cutoff even for all $p < p_c$. This reason is as follows: At each time step, a new node is added and remains as an isolated with the probability $1 - O(1/N)$, which is close to unity as N becomes large. Thus, single-size nodes are accumulated in the system and they are more likely to merge finite-size clusters, reducing the frequency of merging two large clusters. When dy-

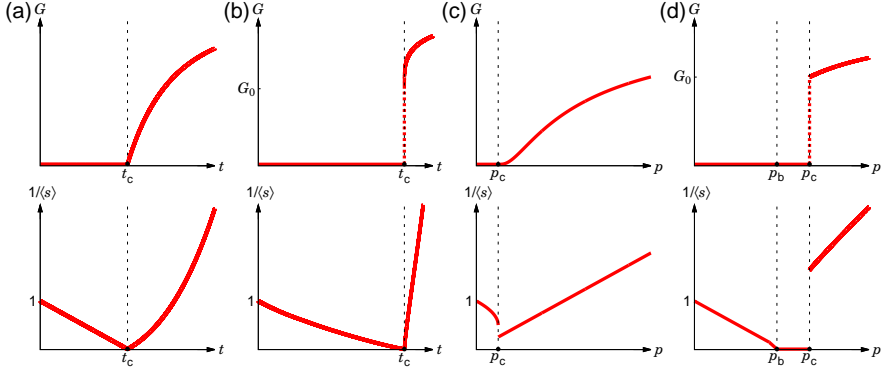


Fig. 3.37: Schematic plots of the order parameter G and the inverse mean cluster size $1/\langle s \rangle$ for the (a) ER, (b) r -ER, (c) GRN, and (d) r -GRN models.

namics reaches a steady state, the cluster merging dynamics self-organizes and forms a power-law behavior of $n_s(p)$. We considered an extreme case that a new node is merged with an existing cluster at each time step with probability p . In this case, $n_s(p)$ is obtained as $\sim s^{-(1+1/p)}$. More generally, as p is increased, more links are added, and the largest cluster becomes larger, and thus the exponent $\tau(p)$ is continuously decreasing. Because the transition point is delayed by the suppression effect, τ can decrease down to two. This eventually leads to a discontinuous PT, because the largest cluster size scales as $N(t)^{1/(\tau-1)}$, where $N(t)$ denotes the system size at a certain time t in steady state, and it reaches up to the extensive size to the system size when $\tau = 2$ regardless t in the steady state.

This tricritical-like behavior at $\tau = 2$ can be seen in the classical polymer aggregation model [52–56]. The cluster aggregation phenomena in a

static system were described via the rate equation,

$$\frac{dn_s(t)}{dt} = \sum_{i+j=s} \frac{w_i n_i}{c(t)} \frac{w_j n_j}{c(t)} - 2 \frac{w_s n_s}{c(t)} \sum_{i=1} \frac{w_i n_i}{c(t)}, \quad (3.15)$$

where $c(t) = \sum_s w_s n_s(t)$. The first term on the R.H.S. represents the aggregation of two clusters of sizes i and j with $i + j = s$, and the second term is for a cluster of size s merging with another cluster of any size. The rate equation reduces to the ER network model when $c(t) = 1$, which occurs when $w_i = i$. A general case, $w_i = i^\omega$, was studied [52–56] long ago. In this case, as ω is smaller, the growth of large clusters is more suppressed. When $1/2 < \omega < 1$, a continuous transition occurs at t_c ; a giant cluster is generated for $t > t_c$. At $t = t_c$, the cluster size distribution follows a power law with exponent $\tau = \omega + (3/2)$. When $0 < \omega \leq 1/2$, a discontinuous transition occurs, and the exponent $\tau = 1 + 2\omega$. The case $\omega = 1/2$, for which $\tau = 2$, is marginal between a second-order and a first-order transition. We remark that another model recent introduced also generates either a continuous or a discontinuous PT by controlling the suppression strength similar to the above case [57]. These two cases are all for static networks. Even though the system type and the underlying mechanism of static and growing networks are different, on the basis of the above result, we could confirm that the discontinuous transition at p_c is induced by the increase of the cluster size heterogeneity across the point with $\tau = 2$.

The BKT transition was found originally in the two-dimensional XY model in thermal systems [58–62]. The origin of the BKT transition in thermal systems is different from that of the percolation model, but there ex-

ist some similarities or dissimilarity in the transition behavior: In the XY model, the singular part of the free energy behaves as $f(t) \sim \exp(-bt^{-1/2})$ with a positive constant b for the reduced temperature $t = (T - T_c)/T_c > 0$, similar to the order parameter $G(p)$ for $p > p_c$ in Eq. (1.1). The correlation function decays in a power-law manner as $\Gamma(r) \sim r^{-\eta(T)}$ for $t < 0$, where $\eta(T) \sim T$ is continuously varying depending on T . This is often called the quasi-long-range order. On the other hand, in a second-order transition $\Gamma(r) \sim r^{-\eta} \exp(-r/\xi)$ for $t < 0$. In this regard, the pure power-law behavior of $\Gamma(r)$ in the infinite-order transition implies $\xi = \infty$ for $t < 0$. Indeed, $\xi = \infty$ for $t < 0$ in the XY model. The continuous varying exponent $\eta(T)$ in the XY model corresponds to the exponent $\tau(p)$ of the cluster size distribution $n_s(p) \sim s^{-\tau(p)}$. In a second-order PT, $n_s \sim s^{-\tau} \exp(-s/s^*)$ for $p < p_c$, where s^* is a characteristic cluster size in the region $p < p_c$. Again, the pure power-law behavior of the infinite-order PT implies that $s^* = \infty$ for $p < p_c$. The susceptibility is obtained using the thermodynamic relation, $\chi \sim \int d^2r \Gamma(r)$. One can find that χ diverges for $\eta < 2$, while it is finite for $\eta > 2$. Because η increases with temperature, χ diverges for $t < 0$ and finite for $t > 0$, where the critical temperature is determined by $\eta = 2$. In percolation, the susceptibility $\chi = \sum_s s^2 n_s$ diverges for $\tau < 3$, while it is finite for $\tau > 3$. For the GRN model, $\tau > 3$ for $p < p_c$. Thus, the susceptibility is finite for $p < p_c$. For $p > p_c$, n_s of finite clusters decays exponentially. Thus the susceptibilities on both sides of a transition point are finite. Even though the order parameter behaves as an infinite-order transition, the susceptibility behavior differently from that of the XY model. On the other hand, for the r -GRN model, the susceptibility diverges in one side and is finite in the other

Table. 3.8: Comparison of the BKT transitions between in thermal systems and in percolations of the growing networks.

Thermal systems	Percolation
$f(t) \sim \exp(-bt^{-1/2}), t = (T - T_c)/T_c$	$G(p) \sim \exp(-a(p - p_c)^{-1/2})$
$\chi \sim \int d^2r \Gamma(r) \sim \int dr r \Gamma(r)$	$\chi = \sum s^2 n_s \approx \int ds s p(s)$
$\Gamma(r) \sim r^{-\eta(T)}$ for $t < 0$	$p(s) = s n_s \sim s^{1-\tau}$ for $p < p_c$
$\eta(T) \sim T < 2$ for $\chi = \infty$	$\tau = \tau(p) < 3$ for $\chi = \infty$
$\xi = \infty$ for $t < 0$	$s^*(p) = \infty$ for $p < p_c$
$\tau(p) > 3$ for $p < p_c$ in the original growing percolation	
$\tau(p) > 2$ for $p < p_c$ in the restricted growing percolation	

side, similar to those of the XY model. These are summarized in Table 3.8.

The BKT transition can occur even in static networks. For instance, the percolation model in one-dimension with $1/r^2$ long-range connections [63] and on hierarchical networks with short-range and long-range connections [64] exhibit BKT infinite-order PTs. As future works, it would be interesting to check whether the diverse phases and phase transitions we obtained occur or not in those static network models when the suppression rule is applied. Moreover, in our study, the suppression rule is applied to large clusters, because the giant cluster size per node is the order parameter in percolation problem. As an extension of our work to thermal systems, it would be interesting to find an essential quantity of thermal BKT systems, for instance, the formation of spin waves or vortices, and see if we can control the BKT transition by the suppression effect. The pattern formation by topological defects in active liquid crystals recently draws considerable attention [65, 66]. Various patterns generated in that system are governed basically by the BKT theory. It would be interesting to note how those patterns are changed when the system is applied by a certain suppression effect.

Chapter 4

Growing scale-free simplicial complexes

4.1 Models and rate equations

This model starts at N_0 isolated nodes and a new node is added every time step t where the total number of nodes is $N(t) = N_0 + t$. Subsequently, $d + 1$ nodes are randomly selected and then they are fully connected with probability p at each time step t . We consider this new formation as a d -simplex.

Now, assuming all finite clusters are trees in the arbitrary phase, the total number of facet degree of nodes belonging to the cluster of size s is written as $((d + 1)/d)(s - 1)$. Then the probability each node, which belongs to cluster of size s_i , is selected to be linked is proportional to $\prod_{i=1}^{d+1} \left(((d + 1)/d)(s_i - 1) + as_i \right)$ where a is the initial attractiveness of nodes. Thus the rate equation of $N_s(p, t)$, defined as the number of cluster of size s , becomes

$$\begin{aligned}
\frac{d(N_s(p, t))}{dt} = & p \left[\sum_{i_1, \dots, i_{d+1}=1}^{\infty} \left(\Pi_{\alpha=1}^{d+1} J_{i_\alpha} N_{i_\alpha} \right) \delta_{\sum_{\alpha=1}^{d+1} i_\alpha, s} \right. \\
& + \sum_{r=1}^{d-1} \sum_{i_1, i_{r+2}, \dots, i_{d+1}=1}^{\infty} \binom{d+1}{r+1} J_{i_1} N_{i_1} \left(\frac{i_1}{N} \right)^r \left(\Pi_{\alpha=r+2}^{d+1} J_{i_\alpha} N_{i_\alpha} \right) \\
& \times \delta_{i_1 + \sum_{\alpha=r+2}^{d+1} i_\alpha, s} - (d+1) J_s N_s - \sum_{r=1}^{d-1} \binom{d+1}{r} J_s N_s \left(\frac{s}{N} \right)^r \Big] + \delta_{1s},
\end{aligned} \tag{4.1}$$

where $J_s(t) \equiv (s((d+1)/d+a) - (d+1)/d)((d+1)pt + aN(t))^{-1}$. In the steady state as time t goes to infinity, Eq. (4.1) is reduced to

$$n_s(p) = p \left[\sum_{i_1, \dots, i_{d+1}=1}^{\infty} \left(\Pi_{\alpha=1}^{d+1} \bar{J}_{i_\alpha} n_{i_\alpha} \right) \delta_{\sum_{\alpha=1}^{d+1} i_\alpha, s} - (d+1) \bar{J}_s n_s \right] + \delta_{1s}, \tag{4.2}$$

where $\bar{J}_s \equiv (s((d+1)/d+a) - (d+1)/d)((d+1)p+a)^{-1}$. The form of \bar{J}_s becomes

$$\bar{J}_s = \frac{s((d+1)/d+a) - (d+1)/d}{(d+1)p+a} = \frac{(d+1)/d+a}{(d+1)p+a} \left(s - \frac{(d+1)/d}{(d+1)/d+a} \right). \tag{4.3}$$

Thus Eq. (4.2) become

$$\begin{aligned}
n_s(p) = & p \left[\left(\frac{(d+1)/d+a}{(d+1)p+a} \right)^{d+1} \sum_{i_1, \dots, i_{d+1}=1}^{\infty} \left(\Pi_{\alpha=1}^{d+1} \left(i_\alpha - \frac{(d+1)/d}{(d+1)/d+a} \right) n_{i_\alpha} \right) \right. \\
& \times \delta_{\sum_{\alpha=1}^{d+1} i_\alpha, s} - (d+1) \left(\frac{(d+1)/d+a}{(d+1)p+a} \right) \left(s - \frac{(d+1)/d}{(d+1)/d+a} \right) n_s \Big] + \delta_{1s},
\end{aligned} \tag{4.4}$$

Now, let us introduce two generating functions $f(x) \equiv \sum_{s=1}^{\infty} s n_s x^s$ and $n(x) \equiv \sum_{s=1}^{\infty} n_s x^s$. Then, one can get

$$f(x) - x + (d+1)p \frac{(d+1)/d + a}{(d+1)p + a} \left(x f'(x) - \frac{(d+1)/d}{(d+1)/d + a} f(x) \right) \\ \times \left\{ 1 - \left(\frac{(d+1)/d + a}{(d+1)p + a} \right)^d \left[f(x) - \frac{(d+1)/d}{(d+1)/d + a} n(x) \right]^d \right\} = 0. \quad (4.5)$$

For convenience, one can define

$$h(x) \equiv \frac{(d+1)/d + a}{(d+1)p + a} \left(f(x) - \frac{(d+1)/d}{(d+1)/d + a} n(x) \right). \quad (4.6)$$

The forms of $f(x)$ and $n(x)$ are thus written as

$$f(x) = x - (d+1)p x \frac{dh(x)}{dx} + p x \frac{dh^{d+1}(x)}{dx} = x + (d+1)p x h'(x) (h^d(x) - 1), \quad (4.7)$$

$$n(x) = \int_0^x dx' \frac{f(x')}{x'} = x - (d+1)p h(x) + p h^{d+1}(x), \quad (4.8)$$

where $f(0)$ and $n(0)$ are $f(0) = 0$ and $n(0) = 0$ leading to $h(0) = 0$. The form of $f(x)$ and $f'(x)$ can be expressed in terms of x , $h(x)$, and $h'(x)$ as follow.

$$f(x) = \frac{(d+1)p + a}{(d+1)/d + a} h(x) + \frac{(d+1)/d}{(d+1)/d + a} (x - (d+1)p h(x) + p h^{d+1}(x)), \quad (4.9)$$

$$f'(x) = \frac{(d+1)p + a}{(d+1)/d + a} h'(x) + \frac{(d+1)/d}{(d+1)/d + a} (1 - (d+1)p h'(x) \\ + (d+1)p h'(x) h^d(x)). \quad (4.10)$$

The giant cluster size G and the second order moment $\langle s \rangle$, defined as $G = 1 - f(1)$ and $\langle s \rangle = f'(1)$, respectively, become thus

$$G = 1 - \frac{(d+1)p+a}{(d+1)/d+a}h(1) - \frac{(d+1)/d}{(d+1)/d+a}(1 - (d+1)ph(1) + ph^{d+1}(1)), \quad (4.11)$$

$$\langle s \rangle = \frac{(d+1)p+a}{(d+1)/d+a}h'(1) + \frac{(d+1)/d}{(d+1)/d+a}(1 - (d+1)ph'(1) + (d+1)ph'(1)h^d(1)). \quad (4.12)$$

Furthermore, substituting Eqs. (4.7) and (4.8) into (4.6), one can get

$$h(x) = \frac{(d+1)/d+a}{(d+1)p+a} \left(x + (d+1)pxh'(x)(h^d(x) - 1) - \frac{(d+1)/d}{(d+1)/d+a} \times (x - (d+1)ph(x) + ph^{d+1}(x)) \right). \quad (4.13)$$

Rearranging for $h'(z)$, Eq. (4.13) is rewritten as

$$\begin{aligned} h'(x) &= \frac{\frac{(d+1)p+a}{(d+1)/d+a}h(x) - x + \frac{(d+1)/d}{(d+1)/d+a}(x - (d+1)ph(x) + ph^{d+1}(x))}{(d+1)px(h^d(x) - 1)} \\ &= \frac{(a - ((d+1)/d)p)h(x) - ax + ((d+1)/d)ph^{d+1}(x)}{(d+1)p((d+1)/d+a)x(h^d(x) - 1)}. \end{aligned} \quad (4.14)$$

When $x = 1$, we classify Eq. (4.14) into those in the normal and percolation phase. In the normal phase, i.e., $f(1) = 1$, assuming all clusters are almost trees, the value of $\sum_{s=1} \left[s((d+1)/d+a) - (d+1)/d \right] N_s(t) = (d+1)pt + aN(t)$ becomes $((d+1)/d+a)f(1) - ((d+1)/d)n(1) = (d+1)p+a$ for large t with leading to $h(1) = 1$. Applying L'Hopital's rule when $x = 1$, one

then get

$$h'(1) = \frac{(a - ((d+1)/d)p)h'(1) - a + ((d+1)^2/d)ph'(1)}{(d+1)p(d+1+ad)h'(1)}. \quad (4.15)$$

Thus,

$$(d+1)p(d+1+ad)h'^2(1) - ((d+1)p+a)h'(1) + a = 0. \quad (4.16)$$

The solutions of Eq. (4.15) are

$$h'(1) = \frac{a + (d+1)p \pm \sqrt{(a + (d+1)p)^2 - 4a(d+1)p(d+1+ad)}}{2(d+1)p(d+1+ad)}. \quad (4.17)$$

The only single solution with minus sign in Eq. (4.17) is valid because $h'(1)$ must be zero when $p = 0$. Moreover, this one is valid for $0 \leq p \leq p_c$, where the discriminant $D(d, a, p)$ of Eq. (4.17), defined as $D(d, a, p) \equiv (a + (d+1)p)^2 - 4a(d+1)p(d+1+ad)$, is non-negative value. Subsequently, the critical point p_c , satisfying $D(d, a, p_c) = 0$, is written as

$$p_c(d, a) = \frac{A(d, a) \pm \sqrt{-4a^2/(d+1)^2 + A^2(d, a)}}{2}, \quad (4.18)$$

where $A(d, a)$ is, for convenience, defined as

$$A(d, a) \equiv -2a\left(\frac{1}{d+1} - 2\right) - 4a^2\left(\frac{1}{d+1} - 1\right). \quad (4.19)$$

Single solution with the minus sign in Eq. (4.18) is only valid owing to the

condition of $\lim_{a \rightarrow \infty} p_c(d, a) = 1/(4d(d+1))$ in the limit of growing random simplicial complexes. Moreover, it is easily confirmed that the asymptotic value of $p_c(d, a)$ goes to 0 for $a \rightarrow 0$. In this phase, the giant cluster size G and the second order moment $\langle s \rangle$ can be obtained from Eqs. (4.11) and (4.12) with the above $h(1)$ and $h'(1)$ for $0 \leq p \leq p_c$.

In the percolation phase with the giant cluster where the value of $f(1)$ is nonzero, The value of $h(1)$ becomes $h(1) = \int_{x=0}^1 h'(x)dx$ and then $h'(1)$ is obtained from $h(1)$ using Eq. (4.14). These values determine the giant cluster size $G = 1 - f(x)$ and $\langle s \rangle = f'(1)$ from Eqs. (4.11) and (4.12).

Finally, the general forms of G and $\langle s \rangle$ are written as

$$G = \begin{cases} 0 & \text{for } p < p_c, \\ 1 - \frac{(d+1)p+a}{(d+1)/d+a} h(1) - \frac{(d+1)/d}{(d+1)/d+a} (1 - (d+1)ph(1) + ph^{d+1}(1)) & \text{for } p \geq p_c, \end{cases} \quad (4.20)$$

$$\langle s \rangle = \begin{cases} \frac{(d+1)p+a}{(d+1)/d+a} \left(\frac{a+(d+1)p - \sqrt{(a+(d+1)p)^2 - 4a(d+1)p(d+1+ad)}}{2(d+1)p(d+1+ad)} + \frac{(d+1)/d}{(d+1)p+a} \right) & \text{for } p < p_c, \\ \frac{(d+1)p+a}{(d+1)/d+a} h'(1) + \frac{(d+1)/d}{(d+1)/d+a} (1 - (d+1)ph'(1) + (d+1)ph'(1)h^d(1)) & \text{for } p \geq p_c, \end{cases} \quad (4.21)$$

where $h(1)$ and $h'(1)$ are $h(1) = \int_0^1 \frac{((d+1)p+a - (d+1)^2 p/d)h(x) - ax + ((d+1)/d)ph^{d+1}(x)}{(d+1)p((d+1)/d+a)x(h^d(x)-1)} dx$

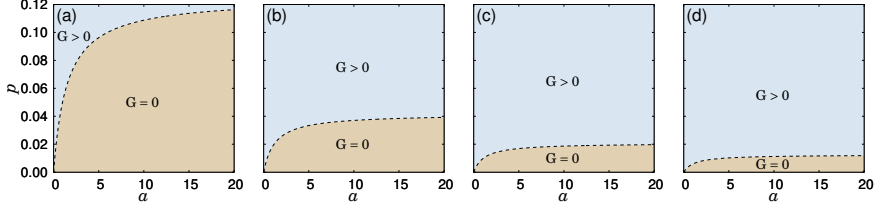


Fig. 4.38: (Color online) Phase diagram of growing scale-free simplicial complexes for (a) $d = 1$ (link), (b) 2 (triangle), (c) 3 (tetrahedron), and (d) 4 (5-cell) in (p, a) plane, where p and a are the probability, which represents the facet density, and the initial attractiveness, respectively. Each solid line represents the critical line satisfying $p = p_c(d, a)$. The giant cluster exists in the region above each critical line for $p > p_c(d, a)$, while it is absent in the region below each critical line for $p < p_c(d, a)$.

$$dx \text{ and } h'(1) = \frac{((d+1)p+a-(d+1)^2p/d)h(1)-a+((d+1)/d)ph^{d+1}(1)}{(d+1)p((d+1)/d+a)(h^d(1)-1)}, \text{ respectively, for}$$

$$p \geq p_c.$$

The left-sided limit of $\langle s \rangle$ is $\frac{(a+(d+1)p_c)}{(d+1)/d+a} \left(\frac{a+(d+1)p_c}{2(d+1)p_c(d+1+ad)} + \frac{(d+1)/d}{(d+1)p_c+a} \right)$ and the right-sided limit of $\langle s \rangle$ is $\frac{(a+(d+1)p_c)}{(d+1)/d+a} \left(\frac{a+(d+1)p_c}{(d+1)p_c(d+1+ad)} + \frac{(d+1)/d}{(d+1)p_c+a} \right)$ since $f(1) = 1$ at $p = p_c$. The difference is always a nonzero value of $\frac{a+(d+1)p_c}{2(d+1)p_c(d+1+ad)}$ for given $a > 0$ and $d \geq 2$. It shows clearly there is discontinuity in Eq. (4.21) at $p = p_c$, resembling the properties of the infinite-order percolation transitions. The phase diagrams are presented in Fig. 4.38 for $d = 1$ (link), 2 (triangle), 3 (tetrahedron), and 4-simplexes (5-cell). When $d = 2$, especially, our results are consistent with those in Ref. [67] for simple model of coauthorship networks [68–71].

4.2 Degree and facet degree distributions

Now, we derive the degree and facet degree distributions in growing scale-free simplicial complexes for the case of d -simplex. The probability, that a

node added at time T is selected to be connected by a new link, is proportional to $k_T(t) + a$, where $k_T(t)$ is the facet degree of the target node at time t and a is the initial attractiveness. The rate equation of $q(k, p, T, t)$, defined as the probability that the node added at time T have the facet degree of k at time t with probability p , is written as

$$\begin{aligned}
& q(k, p, T, t+1) \\
&= p \left[\sum_{l=1}^{d+1} l \binom{d+1}{l} \left(\frac{k-1+a}{(d+1)pt + aN(t)} \right)^l \left(1 - \frac{k-1+a}{(d+1)pt + aN(t)} \right)^{d+1-l} \right. \\
&\quad \times q(k-1, p, T, t) + \left. \left(1 - \frac{k+a}{(d+1)pt + aN(t)} \right)^{d+1} q(k, p, T, t) \right] \\
&\quad + (1-p)q(k, p, T, t), \tag{4.22}
\end{aligned}$$

where the total number of node is $N(t) = N(0) + t$. Summing up Eq. (4.22) over k from 1 to t , one can get

$$\begin{aligned}
& N(t+1)P_f(k, p, t+1) - q(k, p, t+1, t+1) \\
&= p \left[\left(\frac{k-1+a}{pt/N(t) + a/(d+1)} \right) P_f(k-1, p, t) + \left(N(t) - \frac{k+a}{pt/N(t) + a/(d+1)} \right) \right. \\
&\quad \times P_f(k, p, t) \left. \right] + (1-p)N(t)P_f(k, p, t) + O\left(\frac{P_f}{N(t)}\right), \tag{4.23}
\end{aligned}$$

where $P_f(k, p, t) \equiv \sum_{T=0}^t N_0^{\delta_{0r}} q(k, p, T, t)/N(t)$. The initial condition of distribution $q(k, p, t, t)$ is $q(k, p, t, t) = \delta_{0k}$ because facet degree of a new

node is zero. For large t limit, Eq. (4.24) is rewritten as

$$\begin{aligned} \frac{d(N(t)P_f(k, p, t))}{dt} = & \left(\frac{k-1+a}{1+a/((d+1)p)} \right) P_f(k-1, p, t) \\ & - \left(\frac{k+a}{1+a/((d+1)p)} \right) P_f(k, p, t) + \delta_{0k}. \end{aligned} \quad (4.24)$$

Assuming $P_f(k, p, t)$ become time-independent $P_f(k, p)$ in the steady state limit as time goes to infinity, one can easily get

$$\begin{aligned} & \left(1 + \frac{a}{(d+1)p} \right) P_f(k, p) + (k+a)P_f(k, p) - (k-1+a)P_f(k-1, p) \\ & = \left(1 + \frac{a}{(d+1)p} \right) \delta_{0k}. \end{aligned} \quad (4.25)$$

Now, using the generating function $\Phi(z, p)$ of the facet degree distribution $P_f(k, p)$ satisfying $\Phi(z, p) = \sum_{k=0}^{\infty} P_f(k, p)z^k$, one can get

$$z(1-z)\Phi' + a(1-z)\Phi + \left(1 + \frac{a}{(d+1)p} \right) \Phi = \left(1 + \frac{a}{(d+1)p} \right). \quad (4.26)$$

The solution of (4.26) around $z = 0$ is written in terms of the hypergeometric function ${}_2F_1$ as follow.

$$\Phi(z, p) = \frac{1+a/((d+1)p)}{1+a+a/((d+1)p)} {}_2F_1\left[1, a; 2+a+\frac{a}{(d+1)p}; z\right]. \quad (4.27)$$

From the relation ${}_2F_1[a, b; c; z] = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\Gamma[k+a]\Gamma[k+b]\Gamma[c]}{\Gamma[a]\Gamma[b]\Gamma[k+c]} z^k$, the facet degree distribution is finally written as

$$\begin{aligned} P_f(k, p) &= \left(1 + \frac{a}{(d+1)p}\right) \frac{\Gamma[1+a+a/((d+1)p)]}{\Gamma[a]} \frac{\Gamma[k+a]}{\Gamma[2+k+a+a/((d+1)p)]}. \end{aligned} \quad (4.28)$$

For $k \gg 1$, by using the Stirling approximation, Eq. (4.28) becomes

$$P_f(k, p) \simeq \left(1 + \frac{a}{(d+1)p}\right) \frac{\Gamma[1+a+a/((d+1)p)]}{\Gamma[a]} (k+a)^{-(2+\frac{a}{(d+1)p})}, \quad (4.29)$$

leading to the scaling exponent of $\gamma = 2 + a/((d+1)p)$.

Next, for graph degree distributions, we start at the following rate equation:

$$\begin{aligned} q(k, p, T, t+1) &= p \left[\sum_{l=1}^{d+1} l \binom{d+1}{l} \left(\frac{(k-d)/d+a}{(d+1)pt+aN(t)} \right)^l \left(1 - \frac{k/d-1+a}{(d+1)pt+aN(t)} \right)^{d+1-l} \right. \\ &\quad \times q(k-d, p, T, t) + \left. \left(1 - \frac{k/d+a}{(d+1)pt+aN(t)} \right)^{d+1} q(k, p, T, t) \right] \\ &\quad + (1-p)q(k, p, T, t), \end{aligned} \quad (4.30)$$

For degree with respect to the link in graph, the degree distribution $P_g(k, p)$ is written as follow because d links are added to a each node when a d -

simplex is added.

$$\begin{aligned}
P_g(k, p) &= \left(1 + \frac{a}{(d+1)p}\right) \frac{\Gamma[1 + a + a/((d+1)p)]}{\Gamma[a]} \frac{\Gamma[k/d + a]}{\Gamma[2 + k/d + a + a/((d+1)p)]}, \\
\end{aligned} \tag{4.31}$$

leading to the scaling exponent of $\gamma = 2 + a/((d+1)p)$ for large k , which is consistent with the one of facet degree distribution.

In the limit as a goes to infinity without the preferential attachment, two degree distributions of Eqs. (4.28) and (4.31) are reduced to those in growing random simplex complex for the d -simplex.

4.3 Poisson distribution

Rate equation

Now, we consider the d -simplex, whose dimension d follow the Poisson distribution with mean $\lambda \equiv d - 1$, is added to the system every time step. Assuming all finite clusters consist of $\lambda + 1$ on average and they are all trees, the total number of facet degree of nodes belonging to the cluster of size s is written as $((\lambda + 2)/(\lambda + 1))(s - 1)$. Then the probability each node, which belongs to cluster of size s_i , is selected to be linked is proportional to $\prod_{i=1}^{d+1} \left(((\lambda + 2)/(\lambda + 1))(s_i - 1) + as_i \right)$ where a is the initial attractiveness of nodes. Thus the rate equation of $N_s(p, t)$, defined as the number of cluster

of size s , becomes

$$\begin{aligned}
\frac{d(N_s(p, t))}{dt} = & p \sum_{\delta=0}^{\infty} P(\delta; \lambda) \left[\sum_{i_1, \dots, i_{d+1}=1}^{\infty} \left(\Pi_{\alpha=1}^{d+1} J_{i_\alpha} N_{i_\alpha} \right) \delta_{\sum_{\alpha=1}^{d+1} i_\alpha, s} \right. \\
& + \sum_{r=1}^{d-1} \sum_{i_1, i_{r+2}, \dots, i_{d+1}=1}^{\infty} \binom{d+1}{r+1} J_{i_1} N_{i_1} \left(\frac{i_1}{N} \right)^r \left(\Pi_{\alpha=r+2}^{d+1} J_{i_\alpha} N_{i_\alpha} \right) \\
& \times \delta_{i_1 + \sum_{\alpha=r+2}^{d+1} i_\alpha, s} - (d+1) J_s N_s - \sum_{r=1}^{d-1} \binom{d+1}{r} J_s N_s \left(\frac{s}{N} \right)^r \Big] + \delta_{1s},
\end{aligned} \tag{4.32}$$

where $J_s(t) \equiv (s((\lambda+2)/(\lambda+1)+a) - (\lambda+2)/(\lambda+1))((\lambda+2)pt + aN(t))^{-1}$ and $P(\delta; \lambda) \equiv \lambda^\delta e^{-\lambda} / \delta!$ with setting $d = \delta + 1$. In the steady state as time t goes to infinity, Eq. (4.32) is reduced to

$$\begin{aligned}
n_s(p) \\
= & p \sum_{\delta=0}^{\infty} P(\delta; \lambda) \left[\sum_{i_1, \dots, i_{d+1}=1}^{\infty} \left(\Pi_{\alpha=1}^{d+1} \bar{J}_{i_\alpha} n_{i_\alpha} \right) \delta_{\sum_{\alpha=1}^{d+1} i_\alpha, s} - (d+1) \bar{J}_s n_s \right] + \delta_{1s},
\end{aligned} \tag{4.33}$$

where $\bar{J}_s \equiv (s((\lambda+2)/(\lambda+1)+a) - (\lambda+2)/(\lambda+1))((\lambda+2)p + a)^{-1}$.

The form of \bar{J}_s becomes

$$\begin{aligned}
\bar{J}_s &= \frac{s((\lambda+2)/(\lambda+1)+a) - (\lambda+2)/(\lambda+1)}{(\lambda+2)p + a} \\
&= \frac{(\lambda+2)/(\lambda+1) + a}{(\lambda+2)p + a} \left(s - \frac{(\lambda+2)/(\lambda+1)}{(\lambda+2)/(\lambda+1) + a} \right).
\end{aligned} \tag{4.34}$$

Facet and graph degree distributions

Now, we derive the degree and facet degree distributions in growing scale-free simplicial complexes for the case of d -simplex, whose dimension d follow the Poisson distribution with mean $\lambda \equiv d - 1$. The probability, that a node added at time T is selected to be connected by a new link, is proportional to $k_T(t) + a$, where $k_T(t)$ is the facet degree of the target node at time t and a is the initial attractiveness. The rate equation of $q(k, p, T, t)$, defined as the probability that the node added at time T have the facet degree of k at time t with probability p , is written as

$$\begin{aligned}
 q(k, p, T, t+1) = & p \sum_{\delta=0}^{\infty} P(\delta; \lambda) \left[\sum_{l=1}^{d+1} l \binom{d+1}{l} \left(\frac{k-1+a}{(\lambda+2)pt+aN(t)} \right)^l \right. \\
 & \times \left(1 - \frac{k-1+a}{(\lambda+2)pt+aN(t)} \right)^{d+1-l} q(k-1, p, T, t) \\
 & + \left(1 - \frac{k+a}{(\lambda+2)pt+aN(t)} \right)^{d+1} q(k, p, T, t) \Big] \\
 & + (1-p)q(k, p, T, t),
 \end{aligned} \tag{4.35}$$

where the total number of node is $N(t) = N(0) + t$ and $P(\delta; \lambda) \equiv \lambda^\delta e^{-\lambda} / \delta!$

with setting $d = \delta + 1$. By using the Taylor expansion, one can get

$$\begin{aligned}
 q(k, p, T, t+1) = & p \sum_{\delta=0}^{\infty} P(\delta; \lambda) \left[\left(\frac{(d+1)(k-1+a)}{(\lambda+2)pt+aN(t)} \right) q(k-1, p, T, t) \right. \\
 & + \left(1 - \frac{(d+1)(k+a)}{(\lambda+2)pt+aN(t)} \right) q(k, p, T, t) \Big] \\
 & + (1-p)q(k, p, T, t),
 \end{aligned} \tag{4.36}$$

Summing up Eq. (4.36) over k from 1 to t , one can get

$$\begin{aligned}
& N(t+1)P_f(k, p, t+1) - q(k, p, t+1, t+1) \\
&= p \sum_{\delta=0}^{\infty} P(\delta; \lambda) \left[\left(\frac{(d+1)(k-1+a)}{(\lambda+2)pt/N(t)+a} \right) P_f(k-1, p, t) \right. \\
&\quad \left. + \left(N(t) - \frac{(d+1)(k+a)}{(\lambda+2)pt/N(t)+a} \right) P_f(k, p, t) \right] \\
&\quad + (1-p)N(t)P_f(k, p, t) + O\left(\frac{P_f}{N(t)}\right), \tag{4.37}
\end{aligned}$$

where $P_f(k, p, t) \equiv \sum_{T=0}^t N_0^{\delta_{0T}} q(k, p, T, t)/N(t)$. The initial condition of distribution $q(k, p, t, t)$ is $q(k, p, t, t) = \delta_{0k}$ because facet degree of a new node is zero. For large t limit, Eq. (4.38) is rewritten as

$$\begin{aligned}
\frac{d(N(t)P_f(k, p, t))}{dt} &= \sum_{\delta=0}^{\infty} P(\delta; \lambda) \left[\left(\frac{(d+1)(k-1+a)}{\lambda+2+a/p} \right) P_f(k-1, p, t) \right. \\
&\quad \left. - \left(\frac{(d+1)(k+a)}{\lambda+2+a/p} \right) P_f(k, p, t) \right] + \delta_{0k}. \tag{4.38}
\end{aligned}$$

with leading to

$$\begin{aligned}
\frac{d(N(t)P_f(k, p, t))}{dt} &= \left[\left(\frac{k-1+a}{1+a/((\lambda+2)p)} \right) P_f(k-1, p, t) \right. \\
&\quad \left. - \left(\frac{k+a}{1+a/((\lambda+2)p)} \right) P_f(k, p, t) \right] + \delta_{0k}. \tag{4.39}
\end{aligned}$$

Assuming $P_f(k, p, t)$ become time-independent $P_f(k, p)$ in the steady state

limit as time goes to infinity, one can easily get

$$\begin{aligned} & \left(1 + \frac{a}{(\lambda+2)p}\right) P_f(k, p) + (k+a) P_f(k, p) - (k-1+a) P_f(k-1, p) \\ &= \left(1 + \frac{a}{(\lambda+2)p}\right) \delta_{0k}. \end{aligned} \quad (4.40)$$

In the same way as obtaining the Eq. (4.41),

$$\begin{aligned} & P_f(k, p, \lambda) \\ &= \left(1 + \frac{a}{(\lambda+2)p}\right) \frac{\Gamma[1+a+a/((\lambda+2)p)]}{\Gamma[a]} \frac{\Gamma[k+a]}{\Gamma[2+k+a+a/((\lambda+2)p)]}. \end{aligned} \quad (4.41)$$

For $k \gg 1$, by using the Stirling approximation, Eq. (4.41) becomes

$$P_f(k, p) \simeq \left(1 + \frac{a}{(\lambda+2)p}\right) \frac{\Gamma[1+a+a/((\lambda+2)p)]}{\Gamma[a]} (k+a)^{-(2+\frac{a}{(\lambda+2)p})}, \quad (4.42)$$

leading to the scaling exponent of $\gamma = 2 + a/((d_{eff} + 1)p)$, where $d_{eff} \equiv \lambda + 1$.

Next, for graph degree distributions, we start at the following rate equa-

tion:

$$\begin{aligned}
q(k, p, T, t+1) &= p \sum_{\delta=0}^{\infty} P(\delta; \lambda) \left[\sum_{l=1}^{d+1} l \binom{d+1}{l} \left(\frac{(k-d)/d+a}{(\lambda+2)pt+aN(t)} \right)^l \right. \\
&\quad \times \left(1 - \frac{(k-d)/d+a}{(\lambda+2)pt+aN(t)} \right)^{d+1-l} q(k-d, p, T, t) \\
&\quad \left. + \left(1 - \frac{k/d+a}{(\lambda+2)pt+aN(t)} \right)^{d+1} q(k, p, T, t) \right] \\
&\quad + (1-p)q(k, p, T, t), \tag{4.43}
\end{aligned}$$

4.4 Giant cluster size

In this section, we derive the explicit form of G in terms of p near the critical point. First we redefine $h(x)$ as $1 - h(x)$. Then, around $x = 1$ and $p = p_c$, Eq. (4.13) become

$$\varphi'(y)\varphi(y) - \varphi(y) \simeq \frac{a/((d+1)p)}{(1+a/((d+1)p))^2} (y^{1-(d+1+ad)} - y), \tag{4.44}$$

where $y \equiv x^{-1/(d+1+ad)}$ and $\varphi(y)/y \equiv h(x)/(1+a/((d+1)p))$. This approximation is valid only around $x = 1$ and $p = p_c$ for small G . When we set $y = 1$, one can get $\varphi(1) = 0$ for the normal phase, while $\varphi(1) \neq 1$ and $\varphi'(1) = 0$ for the percolation phase. Moreover, $f(x)$ in (4.7)

$$\begin{aligned}
f(x) &= x + (d+1)px \frac{dh(x)}{dx} (1 - (1-h(x))^d) \\
&\simeq x + d(d+1)pxh'(x)h(x), \tag{4.45}
\end{aligned}$$

around $x = 1$ and $p = p_c$. In this regime, the giant cluster size G , defined as $G = 1 - f(1)$, is written as

$$G = 1 - f(1) = \frac{d(d+1)p}{d+1+ad} \left(1 + \frac{a}{(d+1)p}\right)^2 (1 - \varphi(1))\varphi(1) \quad (4.46)$$

Now, let us solve the Eq. (4.44) to find $\varphi(1)$. Near $y = 1$ and $p = p_c$, this equation is reduced to

$$\varphi'(y)\varphi(y) - \varphi(y) \simeq -\left(\frac{1}{4} + \alpha\right)(y-1), \quad (4.47)$$

where α is defined as $\alpha \equiv (a/((d+1)p))(d+1+ad)/(1+a/((d+1)p))^2 - 1/4$. The critical point $p_c(d, a)$ lies on the (d, a) -plane satisfying the condition of $\alpha = 0$. Substituting $(y-1)\psi$ for φ , Eq. (4.47) become

$$(y-1)\psi \frac{d\psi}{dy} = -\left[\left(\psi - \frac{1}{2}\right)^2 + \alpha\right], \quad (4.48)$$

and the solution is written as

$$\ln(C(y-1)) = -\int \frac{\psi d\psi}{(\psi - 1/2)^2 + \alpha}, \quad (4.49)$$

where C is a integration constant. For $\alpha > 0$, the solution of Eq. (4.49) becomes

$$\begin{aligned} & \ln(C(y-1)) \\ &= -\frac{1}{2} \ln\left(\left(\frac{\varphi}{y-1} - \frac{1}{2}\right)^2 + \alpha\right) - \frac{1}{2\sqrt{\alpha}} \left[\frac{\pi}{2} + \arctan\left(\frac{1}{\sqrt{\alpha}}\left(\frac{\varphi}{y-1} - \frac{1}{2}\right)\right)\right], \end{aligned} \quad (4.50)$$

thus,

$$\begin{aligned}
& C(y-1) \\
&= \left(\left(\frac{\varphi}{y-1} - \frac{1}{2} \right)^2 + \alpha \right)^{-1/2} \exp \left[-\frac{1}{2\sqrt{\alpha}} \left[\frac{\pi}{2} + \arctan \left(\frac{1}{\sqrt{\alpha}} \left(\frac{\varphi}{y-1} - \frac{1}{2} \right) \right) \right] \right].
\end{aligned} \tag{4.51}$$

As $y \rightarrow 1$ and $\alpha \rightarrow 0$, satisfying $1 \gg y-1 \gg \exp[-\pi/(2\sqrt{\alpha})]$, the Eq. (4.51) is reduced to

$$\varphi(1) = \frac{1}{C} \exp \left(-\frac{\pi}{2\sqrt{\alpha}} \right). \tag{4.52}$$

Substituting this into Eq. (4.46), one finally arrives at

$$\begin{aligned}
G &\simeq \frac{d(d+1)p}{d+1+ad} \left(1 + \frac{a}{(d+1)p} \right)^2 \frac{1}{C} \exp \left(-\frac{\pi}{2\sqrt{\alpha}} \right), \\
&= \frac{4ad}{C} \exp \left(-\frac{\pi}{2\sqrt{\alpha}} \right), \\
&= \frac{4ad}{C} \exp \left[-\frac{\pi}{2} \left[\frac{a(d+1+ad)}{(d+1)p} \left(1 + \frac{a}{(d+1)p} \right)^{-2} - \frac{1}{4} \right]^{-1/2} \right],
\end{aligned} \tag{4.53}$$

for $p > p_c$ around $p = p_c$ in the percolation phase.

For $\alpha = 0$, the solution of Eq. (4.49) satisfies

$$4 \left[\ln \left(2 + \frac{4\varphi}{1-y} \right) + \frac{2}{2 + \frac{4\varphi}{1-y}} \right] = C - 4 \ln(y-1), \tag{4.54}$$

where C is a integral constant. In other form, this equation becomes

$$\ln \left(\frac{1}{2 + \frac{4\varphi}{1-y}} \right) - \frac{2}{2 + \frac{4\varphi}{1-y}} = \ln[C'(y-1)], \tag{4.55}$$

where C' is a new constant. Thus, one can get

$$\varphi(y) = \frac{y-1}{2} \left(1 + \frac{1}{W[-C''(y-1)]} \right), \quad (4.56)$$

where C'' is a new constant and $W(z)$ is the Lambert function satisfying $W(z) \exp[W(z)] = z$. Using the asymptotic form $W(-z) \simeq \ln z$ for $|z| \ll 1$, this solution around $y = 1$ is reduced to

$$\varphi(y) \simeq \frac{y-1}{2} \left(1 + \frac{1}{\ln[C''(y-1)]} \right). \quad (4.57)$$

Thus, when $y = 1$, one get $\varphi(1) = 0$ and the giant cluster size G becomes zero at the critical point $p = p_c$. This is consistent with the fact that the value of Eq. (4.53) for $p > p_c$ becomes zero as p goes to p_c .

For $\alpha < 0$, the solution of Eq. (4.49) becomes

$$\begin{aligned} & \ln(C(y-1)) \\ &= \left(\frac{1}{4\sqrt{-\alpha}} - \frac{1}{2} \right) \ln \left| \psi - \frac{1}{2} + \sqrt{-\alpha} \right| + \left(\frac{-1}{4\sqrt{-\alpha}} - \frac{1}{2} \right) \ln \left| \psi - \frac{1}{2} - \sqrt{-\alpha} \right|, \end{aligned} \quad (4.58)$$

where C is a constant. By considering the case of the normal phase for $p < p_c$, The value of $\varphi(1)$ should be zero at $y = 1$, because giant cluster is absent.

In this case, a valid solution is written as

$$C(y-1) = \left(\frac{1}{2} - \sqrt{-\alpha} - \frac{\varphi}{y-1} \right)^{\frac{1}{4\sqrt{-\alpha}} - \frac{1}{2}} \left(\frac{1}{2} + \sqrt{-\alpha} - \frac{\varphi}{y-1} \right)^{\frac{-1}{4\sqrt{-\alpha}} - \frac{1}{2}}, \quad (4.59)$$

for $\varphi/(y-1) < 1/2 - \sqrt{-\alpha}$. This equation can be rewritten as

$$\begin{aligned} C'(y-1)^{\frac{4\sqrt{-\alpha}}{1-2\sqrt{-\alpha}}} &= \left(\frac{1}{2} - \sqrt{-\alpha} - \frac{\varphi}{y-1}\right) \left(1 - \frac{2}{2\sqrt{-\alpha}+1} \frac{\varphi}{y-1}\right)^{\frac{-1-2\sqrt{-\alpha}}{1-2\sqrt{-\alpha}}} \\ &\simeq \frac{1}{2} - \sqrt{-\alpha} - \frac{2\varphi^2}{(y-1)^2(1-2\sqrt{-\alpha})}, \end{aligned} \quad (4.60)$$

where C' is a new constant. Around $y = 1$, one can get

$$\frac{\varphi^2}{(y-1)^2} = \left(\frac{1}{2} - \sqrt{-\alpha}\right) \left(\frac{1}{2} - \sqrt{-\alpha} - C'(y-1)^{\frac{4\sqrt{-\alpha}}{1-2\sqrt{-\alpha}}}\right). \quad (4.61)$$

Thus,

$$\varphi(y) = \pm(y-1) \sqrt{\left(\frac{1}{2} - \sqrt{-\alpha}\right) \left(\frac{1}{2} - \sqrt{-\alpha} - C'(y-1)^{\frac{4\sqrt{-\alpha}}{1-2\sqrt{-\alpha}}}\right)}. \quad (4.62)$$

The only plus sign of Eq. (4.62) is valid because the value of φ should be always positive. Finally one get,

$$\varphi(y) \simeq (y-1) \left(\frac{1}{2} - \sqrt{-\alpha} + C''(y-1)^{\frac{4\sqrt{-\alpha}}{1-2\sqrt{-\alpha}}}\right), \quad (4.63)$$

where C'' is a new constant. It confirm that $\varphi(1) = 0$ when $y = 1$, and then the corresponding value of the giant cluster size G becomes zero in the normal phase for $p < p_c$.

In summary, the explicit form of the giant cluster size G , around $p = p_c$,

is written as

$$G = \begin{cases} 0 & \text{for } p < p_c, \\ \frac{4ad}{C} \exp \left[-\frac{\pi}{2} \left[\frac{a(d+1+ad)}{(d+1)p} \left(1 + \frac{a}{(d+1)p} \right)^{-2} - \frac{1}{4} \right]^{-1/2} \right] & \text{for } p \geq p_c, \end{cases} \quad (4.64)$$

where p_c is the critical point satisfying $\alpha = 0$.

4.5 Size distribution

The size distribution is derived by using the inverse Z transform of $f(z) = \sum_{s=1}^{\infty} sn_s x^s$ as follow.

$$sn_s = \oint_c \frac{dz}{2\pi i} f(z) z^{-k-1}, \quad (4.65)$$

where the contour c is a unit circle and counterclockwise closed path encircling the point $z = 0$. Substituting Eq. (4.45) into Eq. (4.65), Eq. (4.65) becomes then

$$n_s = \delta_{1s} + \frac{d(d+1)p}{2} \oint_c \frac{dz}{2\pi i} h^2(z) z^{-k-1}. \quad (4.66)$$

Thus we obtain

$$n_s = \delta_{1s} - \frac{d(d+1)p}{2} (d+1+ad) \left(1 + \frac{a}{(d+1)p} \right)^2 \int_c \frac{dy}{2\pi i} \phi^2 y^{s(d+1+ad)-3}, \quad (4.67)$$

where c' is the new integration contour and $y \equiv z^{-(d+1+ad)}$ and $\varphi(y)/y \equiv h(z)/(1+a/((d+1)p))$. Because there is a singularity at $y = 1$, we change the integration variable y into $1 + u$ for small u . For large s , Eq. (4.67) becomes

$$n_s = \frac{d(d+1)p}{2}(d+1+ad) \left(1 + \frac{a}{(d+1)p}\right)^2 \int_{c''} \frac{du}{2\pi i} \varphi^2(1+u) e^{s(d+1+ad)u}, \quad (4.68)$$

where c'' is the new integral contour.

When $p = p_c$, by using Eq. (4.57), one get the size distribution

$$\begin{aligned} n_s &\simeq \frac{d^2(d+1)p(1+a/((d+1)p))^2}{4(d+1+ad)^2} \frac{1}{s^3 \ln^2 s} \\ &\sim \frac{1}{s^3 \ln^2 s}, \end{aligned} \quad (4.69)$$

for large s at $p = p_c$. The coefficient changes depending on the values of d and a , however, the corresponding critical exponent τ is always 3 at the critical point.

In the same way, when $p < p_c$, by using Eq. (4.63), one get

$$n_s \sim s^{-3-4\sqrt{-\alpha}/(1-2\sqrt{-\alpha})}, \quad (4.70)$$

where α is defined as $\alpha \equiv (a/((d+1)p))(d+1+ad)/(1+a/((d+1)p))^2 - 1/4$. Thus, the critical exponent τ is $\tau = 3 + 4\sqrt{-\alpha}/(1-2\sqrt{-\alpha})$ for $p < p_c$. This is consistent with the direct measures by solving numerically Eq. (4.4) as shown in Fig. 4.39.

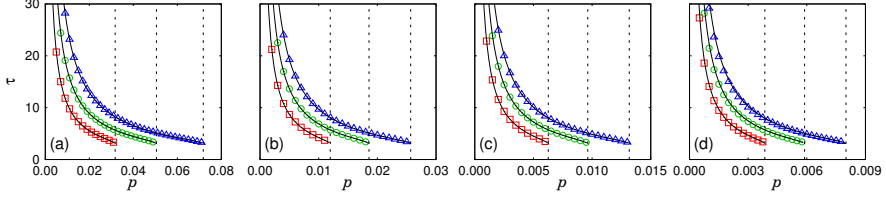


Fig. 4.39: (Color online) Plots of τ versus p in growing scale-free simplicial complexes for (a) $d = 1$ (link), (b) 2 (triangle), (c) 3 (tetrahedron), and (d) 4 (5-cell). Red ($a = 0.5$), green ($a = 1.0$), and blue ($a = 2.0$) symbols represent the results obtained from measures by solving numerically the rate equation of size distribution. Each solid line is the analytic solution $\tau = 3 + 4\sqrt{-\alpha}/(1 - 2\sqrt{-\alpha})$. For each plot, dotted lines from the left represent the critical points for $a = 0.5, 1.0$, and 2.0 .

4.6 The case of $(d + 1)$ -sided polygon

When we consider the case of $(d + 1)$ -sided polygon with $d + 1$ nodes, all properties are the same with those of d -simplex but for the degree distribution about graph. The degree distribution become as follow because only two links are added to a each node when a polygon is added.

$$\begin{aligned}
 P_g(k, p) &= \left(1 + \frac{a}{(d+1)p}\right) \frac{\Gamma[1 + a + a/((d+1)p)]}{\Gamma[a]} \frac{\Gamma[k/2 + a]}{\Gamma[2 + k/2 + a + a/((d+1)p)]}.
 \end{aligned} \tag{4.71}$$

4.7 Analysis

We investigated the percolation phase transition in the growing scale-free simplicial complexes. In our models, 0-simplex is added to the system and a d -simplex is added with probability p . When we select $d + 1$ number of 0 simplexes for d -simplex, we consider that each 0-simplex has a attractive-

ness of their facet degree with initial attractiveness a . In this systems, the degree and facet degree distributions follow the power-law behavior with the same scaling exponent of $\gamma = 2 + a/((d+1)p)$ at p for given d and a . Furthermore, based on the rigorous analytical derivation, we presented that our model exhibits the infinite order phase transition for any d and a . Our findings confirmed that, around the critical point $p = p_c$, giant cluster cluster size increase very smoothly and the second order moment has clear discontinuity. Moreover, our models are reduced to the growing random simplicial complexes when $a \rightarrow \infty$ and the growing scale-free networks when $d = 1$, respectively. Moreover, when $a \rightarrow \infty$, we confirmed degree and facet degree turn to be exponential.

Chapter 5

Machine learning approaches

5.1 Motivation

In appendix, using a unsupervised machine learning (ML) method of principle component analysis (PCA), we investigated the features of percolation variables including the parent node number, occupation number, degree, Adjacency and Laplacian matrix configurations for the percolation model in two-dimensional lattices. As a results we confirmed that the giant cluster sizes are extracted from the parent node number configurations and the occupation probabilities are extracted from the other configurations. Moreover, recently, ML methods have also been successfully used to investigate the percolation properties in two-dimensional lattices [72–74] from occupation number configurations sampled by Monte Carlo (MC) simulations.

However, owing to the difficulty and diversity of network embedding manners, there are no researches about the ML approach to percolation properties in networks. Here, we propose a simple method of network embedding to learn the feature of percolation phase transitions in networks and apply ML techniques to investigating percolation phenomena in networks. We consider typically Erdős-Rényi (ER) and restricted-ER (r -ER) networks. Conventionally, percolation transition type is second order [1, 2, 36, 75] in ER network, whereas the hybrid transition[42] occurs in r -ER network.

5.2 Model and method

Our algorithm comprised three main steps including network embedding, feature learning, and classification (regression) as shown in Fig. 5.40. We first consider ER and r -ER network models. Each networks start at N isolated nodes. All node have number from $i = 1$ to $i = N$ and their parents are themselves at the beginning. When two clusters are merged as the network grows, parents of all nodes in cluster with smaller number of parents node are changed into parents of all node in other cluster before the merger. After the network is grown up, the parent node number of each node in the network is embedded into each site of the one dimensional lattice.

Next, in the feature learning step, the feature of the embedded network information is learned through CNN. Our CNN starts with successive convolutional and pooling layers twice with 8 filters. Kernel size of convolutional layer is 4 and pool size of pooling layer is 4. The values of sites in the last pooling layer are flattened and goes through a fully connected network (FCN). The number of output node of the last layer in this FCN is two for classification of normal and percolation phases by using one-hot encoding, and one for regression of the giant cluster size. For the classification machine, the activation functions of output layer and the other layers in FCN are softmax [76] and the scaled exponential linear units (SELUs) [77] functions, respectively. For the regression machines, the activation functions of all layers in FCN are SELU functions. Loss function is mean squared error and optimizer is Adam [78] for training our machines. The corresponding neural networks in this step are implemented with TensorFlow [79].

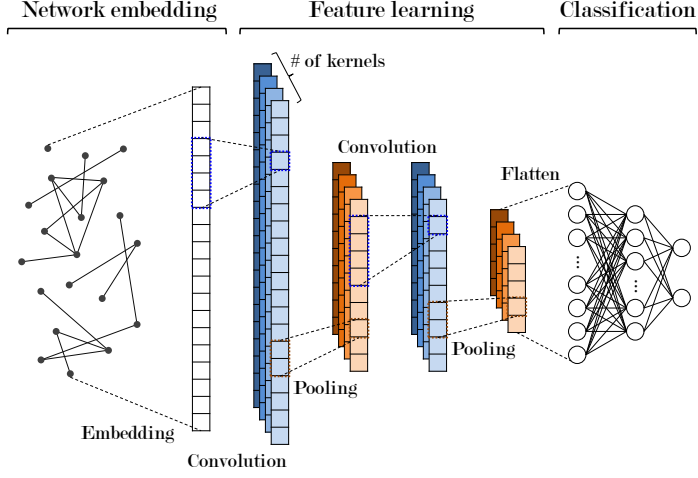


Fig. 5.40: (Color online) Schematic illustration of the machine learning (ML) process in our convolutional neural networks. We first introduce the ER and r -ER networks starting at $N = 20$ isolated nodes and the parents of all nodes are themselves at the beginning. When two clusters are merged as the network grows, parents of all nodes in cluster with smaller number of parents node are changed into parents of all node in other cluster before the merger. For given time t , the parents node numbers of all nodes are embedded into sites on the one dimensional lattice of length size 20. The feature of this embedded information is learned through successive convolutional and pooling layers twice with several filters. kernel size of convolutional layer is 4 and pool size of pooling layer is 4. The values of sites in the last pooling layer are flattened and goes through a fully connected network to classify or regress the information in the network. For the classification machine, the activation functions of output layer and the other layer in FCN are softmax and the scaled exponential linear units (SELUs) functions, respectively. For the regression machines, the activation function of all layer in FCN are SELU functions. Loss function is mean squared error and optimizer is Adam for training our machines.

We perform MC simulations of ER and r -ER networks to collect 1000 configurations of parent node numbers every time resolution of 0.01 in the link density range of $t \in [0, 1.2]$ for the system size $N/10^2 = 2^2, 2^3, \dots, 2^7$. We then remove the information in the range of $[t_c - \Delta t, t_c + \Delta t]$ around the critical, such as $(0.3, 0.7)$ for ER and $(0.55, 0.85)$ for r -ER when $g = 0.8$ case. We also consider two machines of classification and regression and use the parent node number configurations as the input data for training. First, for the classification machine, we set two values (y_1, y_2) of nodes in output layer as $(0, 1)$ for the normal phase in the subcritical region and $(1, 0)$ for the percolation phase in the supercritical region. On the other hand, for the regression machine, we use the corresponding giant cluster size configurations as the target values in output layer for training. We finally collect other 1000 additional configurations of parent node numbers every time resolution of 0.01 in the link density range of $t \in [0, 1.2]$, and 20% of them is for validation and the rest is for test. Our findings from ML are compared to the corresponding MC simulation results by using finite size scaling analysis. Moreover, as test metrics for the regression machines, we use the Pearson correlation coefficient R [80], root mean square error (RMSE), and bias between the giant cluster sizes from our trained machine and MC simulation, where bias is defined as the average of values for ML minus values for MC.

5.3 Results

In this section, we first show the parent node number is the reasonable input data to classify the phases and derive the giant cluster size. Sequentially,

to confirm our classification and regression machines for ER and r -ER networks, we compare the results from ML approach to those from the MC simulation analysis.

5.3.1 Relation between parent node number and giant cluster size

Before using the our parent node number configurations as the training data, we check the relationship between the parent node number and the giant cluster size in ER and r -ER ($g = 0.8$) networks. We consider the total number of nodes N is 12800 with 1000 realizations by using MC simulations for the link density from 0 upto 1.2. Each network starts from N isolated nodes at $t = 0$ where the parents of all nodes are themselves with their initial node numbers. When two clusters are merged following the NZ algorithm, parents of all nodes in cluster, whose parent node number is smaller than other one between two clusters, are changed into parent of other cluster whose parent node number is larger.

Thus the larger the cluster size, the higher the probability that the number of parent nodes is. Moreover, the probability that the number of parent node in the giant cluster is largest among all parents node numbers because it becomes larger and larger as G grows with time t . Fig. 5.41 confirms that the node numbers of parent nodes become larger as the giant cluster sizes G increase with the Spearman correlations coefficients [81] larger than 0.89 for ER and 0.90 for r -ER. We hence expect that the parent node numbers can characterize percolation configurations into normal and percolation phases

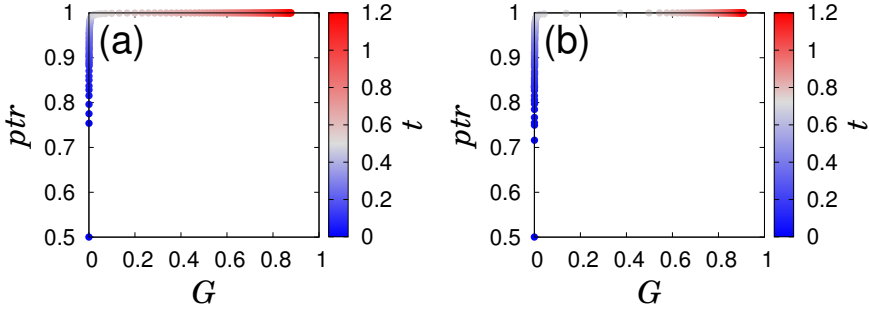


Fig. 5.41: (Color online) For data from Monte Carlo (MC) simulations, plots of the averaged parent node numbers of nodes in the giant cluster versus the giant cluster size for the system size $N/10^2 = 2^7$ with 1000 realizations for (a) ER and (b) r -ER networks. The color of data point represents the link density. The closer link density t to 0 the more blue it becomes, meanwhile the closer its value to 1.2 the more red it becomes. When $t = t_c$, the color is light gray. Spearman correlation coefficients, between the parent node numbers of nodes in the giant cluster and the giant cluster sizes, are 0.89 for ER and 0.90 for r -ER networks. Hence, our results confirm that the probability, that parent node number of the node in the giant cluster is the largest, increases with time t .

clearly, and even directly derive the giant cluster size.

5.3.2 ER network

First, we explored the percolation properties of ER networks where the giant cluster sizes increase continuously at the critical point t_c , exhibiting the second-order phase transition. The critical point t_c and other critical exponents are well known conventionally. We now perform the MC simulation to obtain the giant cluster size G versus t for the system size $N/10^4 = 2^0 - 2^{10}$ over 10^5 realizations. Assuming that the order parameter follow the finite size scaling form of $G(t) = N^{-\beta/\bar{\nu}} f((t - t_c)N^{1/\bar{\nu}})$, we reconfirmed $t_c = 0.50(1)$, $\beta = 1.0(1)$, and $\bar{\nu} = 3.0(1)$. The value of $\bar{\nu}$ is consistent with those obtained from the formula $t_c(N) - t_c(\infty) \sim N^{-1/\bar{\nu}^*}$, where

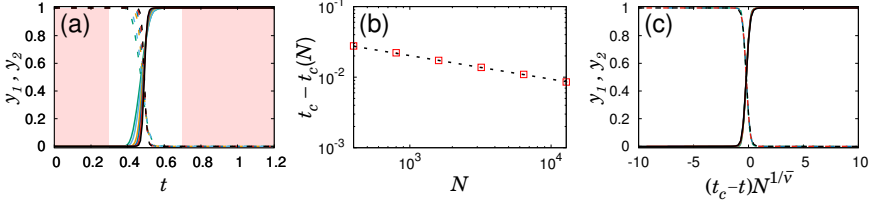


Fig. 5.42: (Color online) For the classification machines of ER networks, plots of (a) output values (y_1, y_2) versus the link density t . Training regions are shaded in orange. (b) Plot of $t_c - t'_c(N)$ versus N , where the asymptotic behavior of $t'_c(N)$ follow $t_c - t'_c(N) \sim N^{-1/\bar{v}'}$ with $t_c = 0.50(1)$ $\bar{v}' = 3.0(1)$. (c) Plot of (y_1, y_2) versus $(t_c - t)N^{1/\bar{v}'}$ and all data are collapsed to a single curve with the estimated t_c and \bar{v}' .

$t_c(N)$ is defined as the point when the dG/dt is the maximum, leading to $\bar{v}^* = 3.0(1)$.

Next, in test results of the classification machines, the intersection points $t'_c(N)$ of output values (y_1, y_2) are estimated for the system size $N/10^2 = 2^2 - 2^7$ in Fig 5.42a. Assuming that the asymptotic behavior of $t'_c(N)$ follows $t_c - t'_c(N) \sim N^{-1/\bar{v}'}$, we obtained $t_c = 0.50(1)$ and $\bar{v}' = 3.0(2)$ as shown in Fig. 5.42b. Using these \bar{v}' , each data point for given system size is well collapsed onto a single curve in Fig. 5.42c. Moreover, the estimated \bar{v}' are consistent with \bar{v} and \bar{v}^* from MC simulation analysis as shown in Table 5.9. It confirm that our classification machines can classify normal and percolation phases in ER networks.

Furthermore, our regression machines are trained to retrieve directly the giant cluster size using configurations of the parent node number. In test results of regression machines, we estimate the Pearson correlation R and root mean square error (RMSE), and bias between the giant cluster sizes from the regression machines and the MC simulations for the sys-

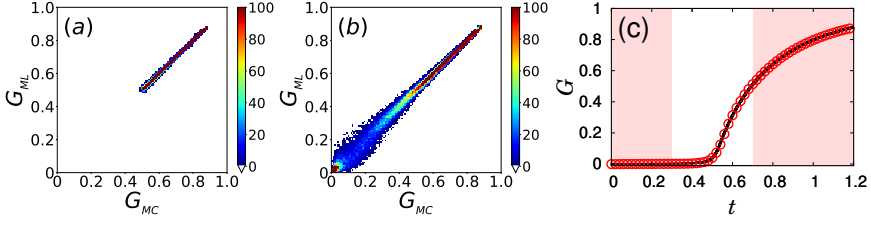


Fig. 5.43: (Color online) Plots of the retrieved values G_{ML} from regression machines versus the giant cluster sizes G_{MC} from Monte Carlo (MC) simulations for (a) training and (b) test results in ER networks for the system size $N/10^2 = 2^7$. In (a), the pixel of (0,0) has been enlarged to make it more visible. (c) The corresponding plot of the giant cluster size (black solid line) from MC simulation versus t , compared to the retrieved value (red open circles) from the regression machine versus t .

tem size $N/10^2 = 2^2 - 2^7$. In Fig. 5.43 and Table 5.10, we confirm that our machine can derive the giant cluster size with R of almost unity and very small RMSE and bias for all system sizes. In the same way as MC simulation, we estimate $t_c(N)$ defined as the point where $dG(t)/dt$ is the maximum. Assuming that the asymptotic behavior of $t_c(N)$ follows the for-

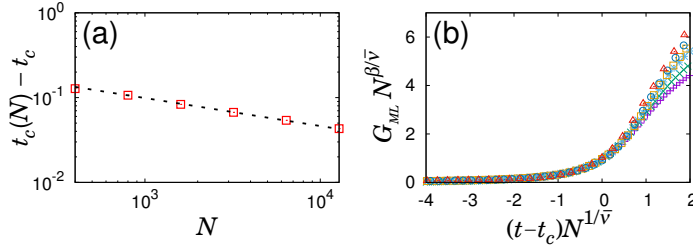


Fig. 5.44: (Color online) (a) Plot of $t_c(N) - t_c$ versus N for $N/10^2 = 2^2 - 2^7$. The value of $t_c(N)$ is defined as the link density where $dG(t)/dt$ is the maximum. Assuming that $t_c(N)$ follow $t_c(N) - t_c \sim N^{-1/\bar{v}^*}$, one can get $t_c = 0.50(1)$ and $\bar{v}^* = 3.0(1)$. From the finite size scaling formula $G(t) = N^{-\beta/\bar{v}} f((t - t_c)N^{1/\bar{v}})$, data points for all system sizes are collapsed to a single curve with $t_c = 0.5(1)$ and $\bar{v} = 3.0(1)$ in (b) plot of $GN^{\beta/\bar{v}}$ versus $(t - t_c)N^{1/\bar{v}}$. All data points above are obtained from our trained regression machines for ER network.

mula $t_c(N) - t_c(\infty) \sim N^{-1/\bar{\nu}^*}$, one can get $t_c = 0.725(2)$ and $\bar{\nu}^* = 3.0(1)$ as shown in Fig. 5.44a. Moreover, by using the finite size scaling form of $G(t) = N^{-\beta/\bar{\nu}} f((t - t_c)N^{1/\bar{\nu}})$, all data points are collapsed to a single curve with $t_c = 0.50(1)$, $\beta = 1.0(1)$ and $\bar{\nu} = 3.0(1)$ in Fig. 5.44b. Our results from ML approaches are consistent with those from MC simulation analysis and summarized in Table 5.9.

5.3.3 r -ER network

In r ER networks, the order parameter jumps discontinuously and then exhibits the second order phase transition, called hybrid PT [42], where the giant cluster sizes $G(t)$ behave as

$$G(t) = \begin{cases} 0 & \text{for } t < t_c, \\ G_0 + (t - t_c)^\beta & \text{for } t \geq t_c. \end{cases} \quad (5.1)$$

In reference [42], the jump size G_0 and the other critical exponents β were presented for $g = 0.1 - 0.9$, continuously varying according to g . Considering $g = 0.8$ case without loss of generality, we use the results of $t_c = 0.725(1)$, $\beta = 0.32(3)$ and $G_0 = 0.29(3)$.

We additionally perform the MC simulation for the system size $N/10^4 = 2^0 - 2^{10}$ over 10^5 realizations. The jump size G_0 is estimated at $G_0(\infty) = 0.29(3)$, satisfying that $G_0(\infty) - G_0(N)$ follow power-law behavior with respect to N . Assuming that the order parameter follow the finite size scaling form of $G(t) - G_0 = N^{-\beta/\bar{\nu}} f((t - t_c)N^{1/\bar{\nu}})$, we reconfirmed $t_c = 0.725(1)$, $G_0 = 0.29(3)$, $\beta = 0.32(3)$, and $\bar{\nu} = 1.8(1)$. In explosive percolation, we

can assume the formula $t_c(\infty) - t_c(N) \sim N^{-1/\bar{\nu}^*}$ holds [82, 83], where $t_c(N)$ is defined as t intercept of the tangent of $G(t)$ at the point where $dG(t)/dt$ is the maximum. From these relations, we get $\bar{\nu}^* = 1.8(1)$ which is consistent with $\bar{\nu} = 1.8(1)$. These results from MC simulations are going to be compared to those from the classification and regression machines.

Now, in test results of the classification machines, the intersection points $t_c'(N)$ of output values (y_1, y_2) are estimated for the system size $N/10^2 = 2^2 - 2^7$ in Fig. 5.45a. Assuming that the asymptotic behavior of $t_c'(N)$ follows $t_c - t_c'(N) \sim N^{-1/\bar{\nu}'}$, we obtained $t_c = 0.725(2)$ and $\bar{\nu}' = 1.8(1)$ as shown in Fig. 5.45b. Using these $\bar{\nu}'$, each data point is well collapsed onto a single curve in Fig. 5.45c. Moreover, the estimated $\bar{\nu}'$ are consistent with $\bar{\nu} = 1.8(1)$ and $\bar{\nu}^* = 1.8(1)$ from MC simulation analysis as shown in Table 5.9. It confirm that our classification machines can classify normal and percolation phases in r -ER networks.

For our regression machines trained to retrieve directly the giant cluster size using configurations of the parent node number, R is almost unity and

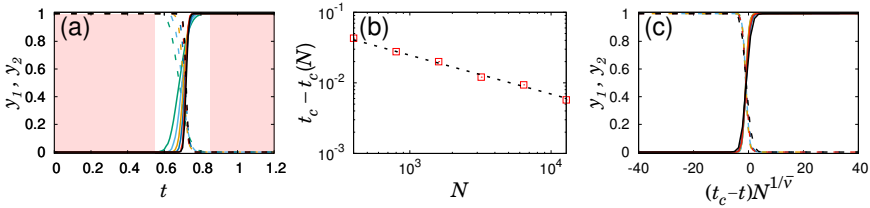


Fig. 5.45: (Color online) For the classification machines of r -ER networks, plots of (a) output values (y_1, y_2) versus the link density t . Training regions are shaded in orange. (b) Plot of $t_c - t_c'(N)$ versus N , where the asymptotic behavior of $t_c'(N)$ follow $t_c - t_c'(N) \sim N^{-1/\bar{\nu}'}$ with $t_c = 0.725(2)$ $\bar{\nu}' = 1.8(1)$. (c) plot of (y_1, y_2) versus $(t_c - t)N^{1/\bar{\nu}'}$ and all data are collapsed to a single curve with the estimated t_c and $\bar{\nu}'$.

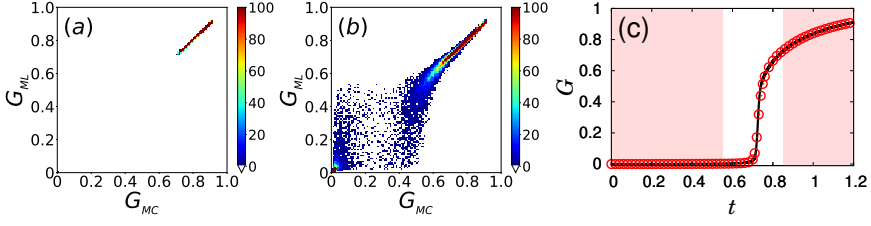


Fig. 5.46: (Color online) Plots of the retrieved value G_{ML} from regression machines versus the giant cluster sizes G_{MC} from MC simulation for (a) training and (b) test results in r -ER networks for the system size $N/10^2 = 2^7$. In (a), the pixel of (0,0) has been enlarged to make it more visible. (c) The corresponding plot of the giant cluster size (black solid line) from MC simulation versus t , compared to the retrieved value (red open circles) from the regression machine versus t .

RMSE is smaller than 10^{-3} with very small bias between the giant cluster sizes from the regression machines and the MC simulations for the system size $N/10^2 = 2^2 - 2^7$ as shown in Fig. 5.46 and Table 5.10. In the same way as MC simulation, Fig.5.47 confirms that the jump size G_0 is estimated at $G_0(\infty) = 0.29(3)$, satisfying that $G_0(\infty) - G_0(N)$ follow power-law behav-

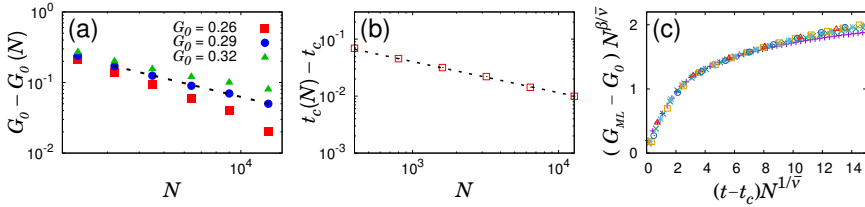


Fig. 5.47: (Color online) (a) Plot of $G_0 - G_0(N)$ versus N for $g = 0.8$ for $N/10^2 = 2^5 - 2^7$, including additional sizes of $N/10^2 = 23, 45$, and 90 . When $G_0 = 0.29(3)$, a power-law decay appears. (b) Plot of $t_c - t_c(N)$ versus N for $N/10^2 = 2^2 - 2^7$. The value of $t_c(N)$ is defined as t intercept of the tangent of $G(t)$ at the point where $dG(t)/dt$ is the maximum. Assuming that $t_c(N)$ follow $t_c - t_c(N) \sim N^{-1/\bar{\nu}^*}$, one can get $t_c = 0.725(2)$ and $\bar{\nu}^* = 1.8(1)$. From the finite size scaling formula $G(t) - G_0 = N^{-\beta/\bar{\nu}} f((t - t_c)N^{1/\bar{\nu}})$, data points for all system sizes are collapsed to a single curve with $t_c = 0.725(2)$, $G_0 = 0.29(3)$, $\beta = 0.32(3)$ and $\bar{\nu} = 1.8(1)$ in (c) plot of $(G - G_0)N^{\beta/\bar{\nu}}$ versus $(t - t_c)N^{1/\bar{\nu}}$. All data above are obtained from our trained regression machines for r -ER.

Table. 5.9: Comparison of the values of the critical point t_c , critical exponent β , jump size G_0 , correlation exponents $\bar{\nu}$ and $\bar{\nu}^*$ obtained from MC simulation and ML approach in Erdős-Rényi (ER) and restricted-ER (r -ER) networks. All values are consistent with each other. The critical exponents $\bar{\nu}$ is defined in the finite size scaling formula for the giant cluster size G in results of MC simulations and regression machines in ER and r -ER. The other exponent $\bar{\nu}^*$ is defined in the formula $|t_c - t_c(N)| \sim N^{-\bar{\nu}^*}$. We remark that $t_c(N)$ is defined as the point where dG/dt is the maximum in ER whereas it is defined as t intercept of the tangent of $G(t)$ at the point where $dG(t)/dt$ is the maximum in r -ER. Additionally, we define $\bar{\nu}'$ in $t_c - t_c'(N) \sim N^{-\bar{\nu}'}$, where $t_c'(N)$ is defined as the intersection point of output values (y_1, y_2) in classification machines in both ER and r -ER networks. The results from ML approaches are consistent with those from MC analysis.

	t_c	β	G_0	$\bar{\nu}$	$\bar{\nu}^*$	$\bar{\nu}'$
ER network						
MC	0.50(1)	1.0(1)	-	3.0(1)	3.0(1)	-
ML	0.50(1)	1.0(1)	-	3.0(1)	3.0(1)	3.0(1)
r -ER network ($g = 0.8$)						
MC	0.725(1)	0.32(3)	0.29(3)	1.8(1)	1.8(1)	-
ML	0.725(2)	0.32(3)	0.29(3)	1.8(1)	1.8(1)	1.8(1)

Table. 5.10: In test data for regression machines, the Pearson correlation coefficient R and root mean square error (RMSE), bias between the giant cluster sizes from regression machines and MC simulations for the system size $N/10^2 = 2^2 - 2^7$ in each ER and r -ER ($g = 0.8$) networks. \bar{G}_{ML} and \bar{G}_{MC} represent the averaged giant cluster size for regression and MC simulation results, respectively.

N	400	800	1600	3200	6400	12800
ER network						
R	0.994	0.995	0.997	0.998	0.998	0.999
RMSE (10^{-4})	363.0	350.7	281.9	236.5	193.1	150.7
Bias (10^{-4})	8.500	-10.61	-22.76	-24.04	-17.40	-7.274
\bar{G}_{ML}	0.373	0.366	0.367	0.360	0.360	0.361
\bar{G}_{MC}	0.372	0.367	0.365	0.362	0.362	0.361
r -ER network ($g = 0.8$)						
R	0.992	0.993	0.995	0.995	0.996	0.997
RMSE (10^{-4})	456.5	460.6	395.4	374.2	331.1	282.2
Bias (10^{-4})	10.55	3.346	10.63	5.952	-8.291	0.832
\bar{G}_{ML}	0.333	0.324	0.321	0.316	0.313	0.313
\bar{G}_{MC}	0.332	0.324	0.320	0.316	0.314	0.313

ior with respect to N . Moreover, from the formula $t_c(\infty) - t_c(N) \sim N^{-1/\bar{\nu}^*}$, where $t_c(N)$ is defined as t intercept of the tangent of $G(t)$ at the point where $dG(t)/dt$ is the maximum, one can get $t_c = 0.725(2)$ and $\bar{\nu}^* = 1.8(1)$. Additionally, by using the finite size scaling form of $G(t) - G_0 = N^{-\beta/\bar{\nu}} f((t - t_c)N^{1/\bar{\nu}})$, we confirm that all data points are collapsed to a single curve with $t_c = 0.725(1)$, $G_0 = 0.29(3)$, $\beta = 0.32(3)$, and $\bar{\nu} = 1.8(1)$. Our results from ML approaches are consistent with those from MC simulation analysis and summarized in Table 5.9.

5.4 Analysis

We have implemented the supervised learning of continuous and hybrid percolation PT of ER and r -ER networks, respectively. We first proposed network embedding method to use the information in networks as the input data of ML. We embed the parent node number of all nodes to the corresponding site in one-dimensional lattice. Next, by using these embedded data, we trained two types of the classification and regression machines and derive the feature of ER and r -ER networks.

Our trained classification machine classified the normal and percolation phases, and gave us the reasonable values of the critical points t_c and the correlation exponents for both ER and r -ER. Furthermore, our trained regression machines could retrieve directly the giant cluster sizes G which is consistent with those of MC simulation results. We thus performed the finite size scaling for the giant cluster size derived from the machines with leading to the critical points t_c , critical exponents β and the correlation exponents $\bar{\nu}$

for ER and r -ER including the jump size G_0 at $t = t_c$. All percolation properties obtained from our ML techniques are consistent with those from MC approach and summarized in Table 5.9.

Those findings show finally ML techniques can be great methods to investigate the percolation properties even in network exhibiting not only continuous PT but also hybrid PT including the discontinuity of the order parameter's emergence.

Chapter 6

Conclusion

We first showed that the critical exponent β decreases algebraically with m in growing networks; however, it decays exponentially in static networks under the local suppression rules. This fact reflects that the suppression effect in growing networks is weaker than that in static networks. Furthermore, we obtained the critical exponents and their tendencies in both growing and static models for arbitrary values of m . The critical exponent β is finite for given m and the order parameter exhibits the second order phase transitions in growing networks under the local suppression effects.

Moreover, we investigated how a BKT PT of growing networks is changed in type when the growth of large clusters in the system is suppressed globally. We introduced the r -GRN model, modified from the GRN model by including the suppression rule. In the r -GRN model, we found that there exist two transition points, p_b and p_c , and three phases. i) In the region $p < p_b$, the order parameter is zero, and the cluster size distribution decays according to a power law without any exponential cutoff and with exponent $\tau(p)$ larger than three. Thus, the mean cluster size is finite. The exponent $\tau(p)$ continuously decreases as p is increased. Accordingly, the region $p < p_b$ is regarded as an infinite-order type critical region. ii) For the region $p_b < p < p_c$, we found that the order parameter is zero, and the cluster size distribution follows a power law without any exponential cut-

off, where the exponent $\tau(p)$ ranges between two and three. Thus, the mean cluster size diverges. This behavior is reminiscent of the critical behavior occurring at the critical point of a second-order transition. Thus, region ii) is regarded as a second-order type critical region. The fact that the mean cluster size diverges, even though the largest cluster has not grown to the extensive size yet, implies that the fluctuations of subextensive-finite clusters diverge preceding to the emergence of the giant cluster of extensive size. Similar behavior occurs in a hierarchical model [12]. iii) At p_c , a discontinuous transition occurs. iv) The region $p > p_c$ is regarded as a noncritical region because the order parameter is finite, and the cluster size distribution decay exponentially. Thus, our model contains the three regimes of the infinite-order, second-order, and first-order transitions. We obtained various properties of the transition behaviors analytically and numerically. We also found that PIN models exhibit the BKT transitions and obtain a similar pattern of PT to those in the r -GRN model. Thus our main results are universal independent of detailed dynamic rules.

Furthermore, we extended the growing networks with edges to the growing scale-free simplicial complexes with d -simplexes. For constructing the growing scale-free simplicial complexes, we consider that each 0-simplex has a attractiveness of their facet degree with a initial attractiveness a . Based on the generating function approach [14, 27–29], we analytically obtained the explicit forms of the giant cluster sizes and size distributions, and confirmed that our models exhibit the infinite-order phase transitions.

Next, we confirm that ML method is successful in exploring the percolation properties in not only lattices but also networks. First we found that

the features of the parent node number configurations are the giant cluster size in two-dimensional percolation system. Based on these results, we performed the supervised learning analysis for ER and r -ER networks. We embedded the parent node number of each node to each corresponding site in one-dimensional lattice. By using these parent node number configurations as the input data of the convolutional neural network (CNN) [84], we developed the classification and regression machines to investigate not only the continuous PT in ER network but also hybrid PT including discontinuous and continuous percolation properties in r -ER network. Finally, we confirmed that our trained classification machines classify the two types of phases with reasonable critical points and the trained regression machines directly retrieve the giant cluster from the parent node number configurations for both ER and r -ER networks, even though the machines were trained with no information around the critical points.

In conclusion, we have solved generally the percolation problems in growing scale-free simplicial complexes covering properties of growing networks, and it can help to understand the properties of growing system. Furthermore, our findings revealed that the machine learning approach is applicable to the percolation issues in networks as well as lattices.

Appendices

Appendix A

r-GRN model

A.1 Rate equations of the *r*-GRN model

Here we recall the rate equations previous derived in Ref. [27]. The cluster number density $n_s(p, t)$ is defined as the number of clusters of size s divided by $N(t)$ at time step t , where p denotes the probability that a link is connected between two selected nodes. We denote the size of the largest cluster in set \mathbf{R} as $S_R(p, t)$. Then the rate equations of $n_s(p, t)$ are as follows:

$$\frac{d(N(t)n_s)}{dt} = p \left[\sum_{i,j=1}^{\infty} \frac{in_i jn_j}{g} \delta_{i+j,s} - sn_s - \frac{sn_s}{g} \right] + \delta_{1s} \quad \text{for } s < S_R, \quad (\text{A.1})$$

$$\frac{d(N(t)n_s)}{dt} = p \left[\sum_{i,j=1}^{\infty} \frac{in_i jn_j}{g} \delta_{i+j,s} - sn_s - \left(1 - \sum_{k=1}^{S_R-1} \frac{kn_k}{g} \right) \right] + \delta_{1s}$$

for $s = S_R$, (A.2)

$$\frac{d(N(t)n_s)}{dt} = p \left[\sum_{j=1}^{\infty} \sum_{i=1}^{S_R-1} \frac{in_i jn_j}{g} \delta_{i+j,s} + \sum_{j=1}^{\infty} \delta_{S_R+j,s} jn_j \left(1 - \sum_{k=1}^{S_R-1} \frac{kn_k}{g} \right) - sn_s \right]$$

for $s > S_R$. (A.3)

On the R.H.S. of equation (A.1) for $s < S_R$, the first term $\sum_{i,j=1}^{\infty} \frac{in_i jn_j}{g} \delta_{i+j,s}$ means the probability that one node is randomly selected in set \mathbf{R} and the other is randomly selected from the entire system, and they are merged and

then generate a cluster of size s . The second and third terms $(1 + 1/g)sn_s$ means the probability that one node is randomly selected from a cluster of size s in set \mathbf{R} regardless of the other node selected from all other nodes in the entire system and vice versa. The last term, δ_{1s} represents the contribution by an incoming isolated node at each time step. In equation (A.2) for $s = S_R$, the first, second, and the last terms are obtained in the same way as in equation (A.1). The third term $1 - \sum_{k=1}^{S_R-1} \frac{kn_k}{g}$ means the probability that one node is randomly selected from the largest cluster of size S_R in set \mathbf{R} regardless of the other node selected from all nodes in the entire system. In equation (A.3) for $s > S_R$, the first term $\sum_{j=1}^{\infty} \delta_{i+j,s} j n_j \sum_{i=1}^{S_R-1} \frac{in_i}{g}$ means the probability that one node is randomly selected from the nodes which do not belong to the cluster of size S_R in set \mathbf{R} and the other node randomly selected in all nodes that generates a cluster of size $s > S_R$. The second term $\sum_{j=1}^{\infty} \delta_{S_R+j,s} j n_j \left(1 - \sum_{i=1}^{S_R-1} \frac{in_i}{g}\right)$ means the probability that one node is randomly selected from the cluster of size S_R in set R and the other node randomly selected from all nodes in the entire system. The third loss term sn_s is obtained in the same way as in equation (A.1) and (A.2). p means the probability that two selected nodes are linked.

In the steady state $t \rightarrow \infty$, $S_R(p, t)$ and $n_s(p, t)$ become independent of t , and they are written as $S_R(p)$ and $n_s(p)$, respectively. Then the L.H.S. of equation (A.1-A.3) become $n_s(p)$ and the R.H.S. of equation (A.1-A.3) are rewritten as follows:

$$n_s = p \left[\sum_{i,j=1}^{\infty} \frac{in_i jn_j}{g} \delta_{i+j,s} - \left(1 + \frac{1}{g}\right) sn_s \right] + \delta_{1s} \quad \text{for } s < S_R, \quad (\text{A.4})$$

$$n_s = p \left[\sum_{i,j=1}^{\infty} \frac{in_i jn_j}{g} \delta_{i+j,s} - sn_s - \left(1 - \sum_{k=1}^{S_R-1} \frac{kn_k}{g}\right) \right] + \delta_{1s}$$

for $s = S_R$, (A.5)

$$n_s = p \left[\sum_{j=1}^{\infty} \sum_{i=1}^{S_R-1} \frac{in_i jn_j}{g} \delta_{i+j,s} + \sum_{j=1}^{\infty} \delta_{S_R+j,s} jn_j \left(1 - \sum_{k=1}^{S_R-1} \frac{kn_k}{g}\right) - sn_s \right]$$

for $s > S_R$. (A.6)

Appendix B

Unsupervised learning of percolation transitions in two-dimensional lattices

B.1 Feature extraction

To collect datasets, we perform MC simulations for two-dimensional bond percolation (2DBP) in L by L of two-dimensional lattices with L^2 nodes. At time $t = 0$, all nodes are isolated with node number from $i = 1$ to $i = N_0$, and their parents are themselves at the beginning. Moreover, a bond can be occupied at one of $2(L - 1)L$ positions with closed boundary conditions every time step t because of the coordination number z is 4. As time t increases, a bond is randomly added to the system at each time with bond occupation probability p defined as $p \equiv t / (2(L - 1)L)$. When two clusters are merged, parents of all nodes in cluster with smaller number of parents node are changed into parents of all node in other cluster before the merger. We grow the system until $p = 1.0$, and collect 2000 configurations of degrees and parent node numbers every resolution of 0.01 in the link density range of $p \in [0, 1.0]$ for the system size $L/10 = 2, 4, 8, 10, 16$, and 20. Half of configurations are used to train the machines with 20% of them for validation, and the rest of them are used for test. We collect additional 2000 configuration of adjacency and Laplacian matrix of size L^2 by L^2 every resolution of 0.01

Table. B.11: The features, extracted by unsupervised learning methods, of percolation variables including site (bond) occupation number, degree, adjacency matrix, Laplacian matrix and parent node number for two-dimensional site (bond) percolation.

	Bond	Site
Bond (site) occupation number	Bond occupation prob.	Site occupation prob.
Degree	Bond occupation prob.	Activated bond density
Adjacency matrix	Bond occupation prob.	Activated bond density
Laplacian matrix	Bond occupation prob.	Activated bond density
Parent node number	Giant cluster size	Giant cluster size

in the link density range of $p \in [0, 1.0]$ for the system size $L/10 = 10$. Similarly, we collect datasets of two-dimensional site percolation (2DSP) for site occupation number and parent node number configurations by performing MC simulations. The foregoing two dimensional variables of configurations are used in both input and output layers in PCA.

After performing PCA, linearity appear between the first principal component y_1 and the giant cluster sizes when the input datasets are the parent node number configurations, whereas, linearity appear between the first principal components y_1 and the bond occupation probabilities when the input datasets are the others.

In the same way as the case of 2DBP, as a result of PCA for 2DSP, linearity appear between the first principal component y_1 and the giant cluster sizes when the input datasets are the parent node number configurations, whereas, linearity appear between the first principal components y_1 and the site occupation probabilities when the input datasets are the site occupation number configurations.

Bibliography

- [1] D. Stauffer, Physics Reports **54**, 1 (1979).
- [2] D. Stauffer and A. Aharony, *Introduction To Percolation Theory : Second Edition* (Taylor & Francis, 2018).
- [3] P. J. Flory, Journal of the American Chemical Society **63**, 3083 (1941).
- [4] S. R. Broadbent and J. M. Hammersley, Mathematical Proceedings of the Cambridge Philosophical Society **53**, 629 (1957).
- [5] E. M. Hendriks, M. H. Ernst, and R. M. Ziff, Journal of Statistical Physics **31**, 519 (1983).
- [6] F. Leyvraz, Physics Reports **383**, 95 (2003).
- [7] P. Grassberger, Mathematical Biosciences **63**, 157 (1983).
- [8] D. Mollison, Journal of the Royal Statistical Society. Series B (Methodological) **39**, 283 (1977).
- [9] J. D. Murray, *Mathematical Biology*, 2nd ed., Biomathematics (Springer-Verlag, Berlin Heidelberg, 1993).
- [10] J. L. Cardy and P. Grassberger, Journal of Physics A: Mathematical and General **18**, L267 (1985).
- [11] M. Walther, D. G. Cooke, C. Sherstan, M. Hajar, M. R. Freeman, and F. A. Hegmann, Physical Review B **76**, 125408 (2007).
- [12] S. Boettcher, V. Singh, and R. M. Ziff, Nature Communications **3**, 787 (2012).
- [13] J. Kim, P. L. Krapivsky, B. Kahng, and S. Redner, Physical Review E **66**, 055101 (2002).

- [14] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz, *Physical Review E* **64**, 041902 (2001).
- [15] J. Chalupa, P. L. Leath, and G. R. Reich, *Journal of Physics C: Solid State Physics* **12**, L31 (1979).
- [16] D. Achlioptas, R. M. D'Souza, and J. Spencer, *Science* **323**, 1453 (2009).
- [17] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, *Nature* **464**, 1025 (2010).
- [18] S.-W. Son, P. Grassberger and M. Paczuski, *Physical Review Letters* **107**, 195702 (2011).
- [19] R. A. da Costa, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Physical Review Letters* **105**, 255701 (2010).
- [20] O. Riordan and L. Warnke, *Science* **333**, 322 (2011).
- [21] Y. S. Cho, S. Hwang, H. J. Herrmann, and B. Kahng, *Science* **339**, 1185 (2013).
- [22] P. Grassberger, C. Christensen, G. Bizhani, S.-W. Son, and M. Paczuski, *Physical Review Letters* **106**, 225701 (2011).
- [23] R. K. Pan, M. Kivelä, J. Saramäki, K. Kaski and J. Kertész, *Physical Review E* **83**, 046112 (2011).
- [24] S. D. Yi, W. S. Jo, B. J. Kim, and S.-W. Son, *Europhysics Letters* **103**, 26004 (2013).
- [25] W. S. Jo, S. D. Yi, B. J. Kim, and S.-W. Son, *Journal of the Korean Physical Society* **65**, 1985 (2014).
- [26] S. M. Oh, S.-W. Son, and B. Kahng, *Physical Review E* **93**, 032316 (2016).

- [27] S. M. Oh, S.-W. Son, and B. Kahng, *Physical Review E* **98**, 060301 (2018).
- [28] S. M. Oh, S.-W. Son, and B. Kahng, *Journal of Statistical Mechanics: Theory and Experiment* **2019**, 083502 (2019).
- [29] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Physical Review E* **64**, 066110 (2001).
- [30] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, *Physical Review Letters* **85**, 4626 (2000).
- [31] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Physical Review Letters* **85**, 4633 (2000).
- [32] S. N. Dorogovtsev and J. F. F. Mendes, *Physical Review E* **63**, 056125 (2001).
- [33] R. Albert and A.-L. Barabási, *Reviews of Modern Physics* **74**, 47 (2002), publisher: American Physical Society.
- [34] R. Albert, H. Jeong, and A.-L. Barabási, *Nature* **401**, 130 (1999).
- [35] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [36] P. Erdos and A. Rényi, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 17 (1960).
- [37] B. Bollobás, *Random Graphs* (Academic Press, New York, 1985).
- [38] R. A. da Costa, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Physical Review E* **89**, 042148 (2014).
- [39] R. A. da Costa, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Physical Review E* **90**, 022145 (2014).
- [40] S. M. Oh, S.-W. Son, and B. Kahng, arXiv:2012.10749 [physics] (2020).

- [41] K. Panagiotou, R. Spöhel, A. Steger, and H. Thomas, Electronic Notes in Discrete Mathematics The Sixth European Conference on Combinatorics, Graph Theory and Applications, EuroComb 2011, **38**, 699 (2011).
- [42] Y. S. Cho, J. S. Lee, H. J. Herrmann, and B. Kahng, Physical Review Letters **116**, 025701 (2016).
- [43] A. Arenas, A. Fernández, S. Fortunato, and S. Gómez, Journal of Physics A: Mathematical and Theoretical **41**, 224001 (2008).
- [44] A. R. Benson, D. F. Gleich, and J. Leskovec, Science **353**, 163 (2016).
- [45] R. Lambiotte, M. Rosvall, and I. Scholtes, Nature Physics **15**, 313 (2019).
- [46] G. Bianconi, Europhysics Letters **111**, 56001 (2015).
- [47] C. Giusti, R. Ghrist, and D. S. Bassett, Journal of Computational Neuroscience **41**, 1 (2016).
- [48] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, 2010).
- [49] K. Choi, D. Lee, Y. S. Cho, J. C. Thiele, H. J. Herrmann, and B. Kahng, Physical Review E **96**, 042148 (2017).
- [50] R. V. Solé and R. Pastor-Satorras, Handbook of Graphs and Networks , 145 (2002).
- [51] R. V. Solé, R. Pastor-Satorras, E. Smith, and T. B. Kepler, Advances in Complex Systems **05**, 43 (2002).
- [52] R. M. Ziff, E. M. Hendriks, and M. H. Ernst, Physical Review Letters **49**, 593 (1982).
- [53] F. Leyvraz and H. R. Tschudi, Journal of Physics A: Mathematical and General **14**, 3389 (1981).

- [54] Y. S. Cho, B. Kahng, and D. Kim, Physical Review E **81**, 030103 (2010).
- [55] S.-W. Son, G. Bizhani, C. Christensen, P. Grassberger, and M. Paczuski, Europhysics Letters **95**, 58007 (2011).
- [56] S.-W. Son, C. Christensen, G. Bizhani, P. Grassberger, and M. Paczuski, Physical Review E **84**, 040102 (2011).
- [57] G. Bianconi, Physical Review E **97**, 022314 (2018).
- [58] V. L. Berezinsky, Soviet Physics JETP **32**, 493 (1971).
- [59] V. L. Berezinsky, Soviet Physics JETP **34**, 610 (1972).
- [60] J. M. Kosterlitz and D. J. Thouless, Journal of Physics C: Solid State Physics **5**, L124 (1972).
- [61] J. M. Kosterlitz and D. J. Thouless, Journal of Physics C: Solid State Physics **6**, 1181 (1973).
- [62] J. M. Kosterlitz, Journal of Physics C: Solid State Physics **7**, 1046 (1974).
- [63] P. Grassberger, Journal of Statistical Mechanics: Theory and Experiment **2013**, P04004 (2013).
- [64] A. N. Berker, M. Hinczewski, and R. R. Netz, Physical Review E **80**, 041118 (2009).
- [65] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha, Reviews of Modern Physics **85**, 1143 (2013).
- [66] X. Tang and J. V. Selinger, Soft Matter **13**, 5481 (2017).
- [67] Y. Lee, J. Lee, S. M. Oh, D. Lee, and B. Kahng, arXiv:2010.12224 [cond-mat, physics:physics] (2021).

- [68] M. E. J. Newman, Proceedings of the National Academy of Sciences **98**, 404 (2001).
- [69] S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, A. Vespignani, L. Waltman, D. Wang, and A.-L. Barabási, Science **359**, eaao0185 (2018).
- [70] D. Lee, K.-I. Goh, B. Kahng, and D. Kim, Physical Review E **82**, 026112 (2010).
- [71] X. Kong, Y. Shi, S. Yu, J. Liu, and F. Xia, Journal of Network and Computer Applications **132**, 86 (2019).
- [72] W. Zhang, J. Liu, and T.-C. Wei, Physical Review E **99**, 032142 (2019).
- [73] W. Yu and P. Lyu, Physica A: Statistical Mechanics and its Applications **559**, 125065 (2020).
- [74] J. Shen, W. Li, S. Deng, and T. Zhang, arXiv:2101.06392 [cond-mat, physics:physics] (2021).
- [75] K. Christensen and N. R. Moloney, *Complexity and Criticality* (World Scientific Publishing Company, 2005).
- [76] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, Vol. 1 (MIT press Cambridge, 2016).
- [77] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17 (Curran Associates Inc., Red Hook, NY, USA, 2017) pp. 972–981.
- [78] D. P. Kingma and J. Ba, arXiv:1412.6980 [cs] (2014).
- [79] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow,

- A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, arXiv:1603.04467 [cs] (2016).
- [80] K. Pearson, Proceedings of the Royal Society of London **60**, 489 (1897).
- [81] L. Myers and M. J. Sirois, in *Encyclopedia of Statistical Sciences* (American Cancer Society, 2004).
- [82] Y. S. Cho, S.-W. Kim, J. D. Noh, B. Kahng, and D. Kim, Physical Review E **82**, 042102 (2010).
- [83] D. Lee, M. Jo, and B. Kahng, Physical Review E **94**, 062307 (2016).
- [84] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Proceedings of the IEEE **86**, 2278 (1998).

초 록

여과 이론은 주어진 계에서 대형 클러스터의 형성을 설명한다. 모든 클러스터들이 무작위로 연결되며 성장할 경우 계의 구조에 관계없이 임계점에서 연속적으로 대형 클러스터가 나타난다. 이러한 무작위성만으로는 폭발적 여과 상전이의 특성을 설명할 수 없기 때문에, 지난 십여 년간 여과 상전이의 종류를 폭발적 여과 상전이로 바꾸기 위하여 작은 클러스터의 성장을 촉진하고 큰 클러스터의 성장은 방해하는 다양한 국소적 억제 규칙이 제시되어왔다. 전체 노드의 개수가 일정한 정적 네트워크에서 큰 클러스터의 성장이 국소적으로 억제되면, 대형 클러스터가 매우 폭발적으로 나타나지만 그 과정이 불연속적이지 않고 연속적이다. 이때 모든 노드들의 정보를 전역적으로 이용하여 큰 클러스터의 성장을 억제하면, 대형 클러스터가 임계점에서 완전히 불연속적으로 나타나게 된다. 하지만 실제 세상에서는 시스템이 성장하는 경우가 많으므로 본 연구에서는 시간에 따라 노드가 증가하는 성장 네트워크에 억제 규칙이 적용되는 경우를 고려하였다. 다양한 종류의 모델에 대하여 클러스터 크기 분포에 대한 비율 방정식을 세우고 이로부터 해석적으로 상전이의 종류를 조사해본 결과, 국소적 억제 조건하에서는 2차 상전이의 특징이 나타나고 전역적 억제 조건 하에서는 1차 상전이의 특징이 나타났다. 이때 성장 네트워크에 국소적 억제 규칙이 적용되었을 경우 기존에 존재하던 노드들이 링크로 연결될 후보군으로 선택될 확률이 높기 때문에 정적 네트워크에서보다 대형 클러스터가 좀 더 부드럽게 나타남을 확인하였다. 또한 전역적인 억제 규칙이 적용되었을 경우에는 무한차 상전이에서의 특징과 유사하게 대형 클러스터가 생기기 전 클러스터 크기 분포가 지수함수를 따르는 임계 영역이

존재함을 확인할 수 있었다. 다음으로, 더 나아가 실제 세계의 도수 분포가 척도가 없고 두 개체 이상이 서로 상호작용한다는 특징에 기반하여 성장 네트워크를 척도 없는 성장 단체 복합체로 확장하였다. 척도 없는 성장 단체 복합체가 억제 규칙 하에서 성장할 때 나타나는 상전이의 종류는 네트워크에서의 경우와 동일함을 확인하였다. 마지막으로, 지도 및 비지도 기계학습 방법을 이용하여 여과 상전이를 기술하는 물리 변수들의 특성을 조사하였다. 그 결과 부모 노드 숫자와 점유 숫자 구성 집합들의 대표 특성이 각각 대형 클러스터의 크기와 점유 확률임을 알 수 있었고, 이러한 성질을 이용해서 네트워크에서의 여과상전이 또한 성공적으로 지도학습하였다. 결과적으로, 본 연구는 일반적인 성장하는 시스템을 해석적으로 이해하는 데 도움을 줄 수 있으며, 기계학습 방법이 격자뿐만 아니라 네트워크에도 성공적으로 적용될 수 있음을 보여준다.

주요어 : 여과 상전이, 폭발적 여과 상전이, 불연속 여과 상전이, 비율 방정식, 생성 함수, 유한 크기 축적 이론, 척도 없는 네트워크, 단체 복합체, 기계 학습

학번 : 2016-30097