OPEN

Statistical Thermodynamics of Irreversible Aggregation: The Sol-Gel Transition

Themis Matsoukas

Department of Chemical Engineering, Pennsylvania State University, University Park, PA 16802.

Binary aggregation is known to lead, under certain kinetic rules, to the coexistence of two populations, one consisting of finite-size clusters (sol), and one that contains a single cluster that carries a finite fraction of the total mass (giant component or gel). The sol-gel transition is commonly discussed as a phase transition by qualitative analogy to vapor condensation. Here we show that the connection to thermodynamic phase transition is rigorous. We develop the statistical thermodynamics of irreversible binary aggregation in discrete finite systems, obtain the partition function for arbitrary kernel, and show that the emergence of the gel cluster has all the hallmarks of a phase transition, including an unstable van der Waals loop. We demonstrate the theory by presenting the complete pre- and post-gel solution for aggregation with the product kernel.

A ubiquitous problem in the physics of dispersed systems is binary aggregation: two clusters i and j merge with probability proportional to the aggregation kernel k_{ij}, a function that characterizes the physics of $i - j$ encounters. This process describes many physical phenomena over a length scales that encompass molecular systems, social networks, and stars[1,2]. Under certain kinetic rate laws that preferentially promote the merging of large clusters, this process produces a remarkable behavior, a *phase transition* manifested in the emergence of a single element that contains a finite fraction of the members of the population. This transition is seen in experimental systems (gelation) as well in dynamic stochastic models, most notably percolation[3]. The standard mathematical tool is Smoluchowski's coagulation equation, developed nearly 100 years ago[4,5]. It is an integral-differential equation that governs the evolution of the mean cluster size distribution of an infinite system whose total mass is fixed. The Smoluchowski equation forms the basis for the quantitative study of colloids and polymers, atmospheric aerosols, animal populations, and dispersed populations in general[2,6,7] and its mathematical behavior has been studied extensively[1,3,8].

The product kernel $k_{ij} = ij$ is a classical example of a rate law that produces a giant cluster within finite aggregation time[1]. It is a model for polymer gelation (polymerization of f-functional monomers in the limit $f \to \infty$[2,9]) and for percolation on random graphs[10,11]. As such, it serves as the standard analytic model for the study of the giant component. In the presence of the giant cluster ("gel phase") the Smoluchowski equation breaks down: the second moment of the size distribution diverges at the gel point (the divergence defines the gel point), and past this point the first moment decays, i.e., mass is not conserved. To restore consistency one assumes the presence of a gel phase (not predicted by the Smoluchowski equation itself) and introduces an additional assumption as to whether the finite-size clusters ("sol phase") interact with the gel (Flory model) or not (Stockmayer model)[12]. These heuristic assumptions lead to different solutions each.

The limitation of the Smoluchowski equation arises from the fact that it reduces an inherently discrete finite stochastic process into a single metric, the mean cluster size distribution. The need for stochastic treatments has been raised in the literature. The discrete finite problem was originally formulated by Marcus[14] and studied in detail by Lushnikov[15–17]. Spouge[18] and Hendriks et al.[19] used a combinatorial approach to obtain solutions for special, non-gelling cases. Other discrete, branching solutions have been implemented in closely related problems in aggregation and fragmentation[20,21], stochastic structures[3], biosciences[22] and astrophysics[23].

These treatments offer results for a number of special cases, but are limited to non-gelling conditions and do not provide a coherent framework for the general case. Lushnikov's method deserves special mention because it is the only one that is capable of obtaining solutions in the post gel region[16]. However, it is mathematically dense and remains opaque to *physical* interpretation. The central question remains unresolved: what is the nature of sol-gel transition and how it relates to familiar phase transitions? We are now in position to answer this question.

Recently we developed a statistical thermodynamic formalism to describe the behavior of generic populations[13]. Here we apply this theory to irreversible aggregation, develop the thermodynamics the discrete finite domain, and obtain the solution to the product kernel.

The Cluster Ensemble

We cast the problem in the theory of the cluster ensemble[13], which we briefly summarize here. We consider a population of M individuals ("monomers") that form N clusters and construct the microcanonical ensemble of all possible distributions $\mathbf{n} = (n_1, n_2, \cdots)$, where n_i is the number of clusters with i monomers. All distributions of the (M, N) ensemble satisfy the two constraints

$$\sum n_i = N, \quad \sum i n_i = M. \tag{1}$$

When two clusters in distribution $\mathbf{n}$ merge, the outcome is a new distribution in the ensemble $(M, N-1)$ of the next generation. We formally define generation $g = M - N + 1$ such that $g = 1$ refers to fully dispersed monomers and $g = M$ to a fully gelled state. These parent-offspring relationships produce a directed graph that represents the phase space of discrete binary aggregation (Fig. 1). Following[13], we express the probability of distribution $\mathbf{n}$ in the (M, N) ensemble as

$$P(\mathbf{n}) = \mathbf{n}! W(\mathbf{n}) / \Omega_{M,N}, \tag{2}$$

where $\mathbf{n}! = N!/n_1! n_2! \cdots$ is the multinomial coefficient, $\Omega_{M,N}$ is the partition function, and $W(\mathbf{n})$ is the bias of distribution, a functional of $\mathbf{n}$ that is determined by the physics of the problem, here, by k_{ij}. The most probable distribution (MPD) in the thermodynamic limit maximizes $\Omega_{M,N}$, and is given by

$$\frac{\tilde{n}_i}{N} = \tilde{w}_i \frac{e^{-\beta i}}{q}, \tag{3}$$

where β, $\log q$ and $\log \tilde{w}_i$ are given by the partial derivatives,

$$\beta = \left(\frac{\partial \log \Omega_{M,N}}{\partial M} \right)_N, \tag{4}$$

$$\log q = \left(\frac{\partial \log \Omega_{M,N}}{\partial N} \right)_M, \tag{5}$$

$$\log \tilde{w}_i = \left(\frac{\partial \log W(\tilde{\mathbf{n}})}{\partial \tilde{n}_i} \right)_{\tilde{n}_j}. \tag{6}$$

These general results, established in Ref. 13, describe the thermodynamic state of a generic population in terms of of the selection bias $W(\mathbf{n})$. We now proceed to derive the selection bias and partition function for the specific problem of interest, the discrete binary aggregation process depicted in Fig. 1.

Aggregation in Discrete Finite Systems

The probability $P(\mathbf{n})$ of distribution $\mathbf{n}$ propagates from parents to offsprings according to equation

$$P(\mathbf{n}) = \sum_{\mathbf{n}' \to \mathbf{n}} P(\mathbf{n}') P_{\mathbf{n}' \to \mathbf{n}}, \tag{7}$$

where $\mathbf{n}'$ is the parent that produces $\mathbf{n}$ by merging cluster sizes $i - j$ and j, and $P_{\mathbf{n}' \to \mathbf{n}}$ is the transition probability for the aggregation event $(i) + (j) \to (i + j)$. The parent distribution $\mathbf{n}'$ is transformed into the offspring distribution $\mathbf{n}$ via the aggregation of cluster masses i and j. The probability for this process is proportional to the number of (unordered) $i - j$ pairs and the aggregation kernel k_{ij}:

$$P_{\mathbf{n}' \to \mathbf{n}} \propto \frac{n'_{i-j} \left(n'_j - \delta_{i-j,j} \right)}{1 + \delta_{i-j,j}} k_{i-j,j}, \tag{8}$$

where δ_{ij} is Kronecker's delta. Normalizing overall all pairs in $\mathbf{n}'$, the transition probability becomes

$$P_{\mathbf{n}' \to \mathbf{n}} = \frac{2}{N(N+1) \bar{k}_{\mathbf{n}'}} \frac{n'_{i-j} \left(n'_j - \delta_{i-j,j} \right)}{1 + \delta_{i-j,j}} k_{i-j,j}, \tag{9}$$

where $\bar{k}_{\mathbf{n}'}$ is the mean aggregation kernel among all pairs in parent $\mathbf{n}'$ (see Supplementary Information for details),
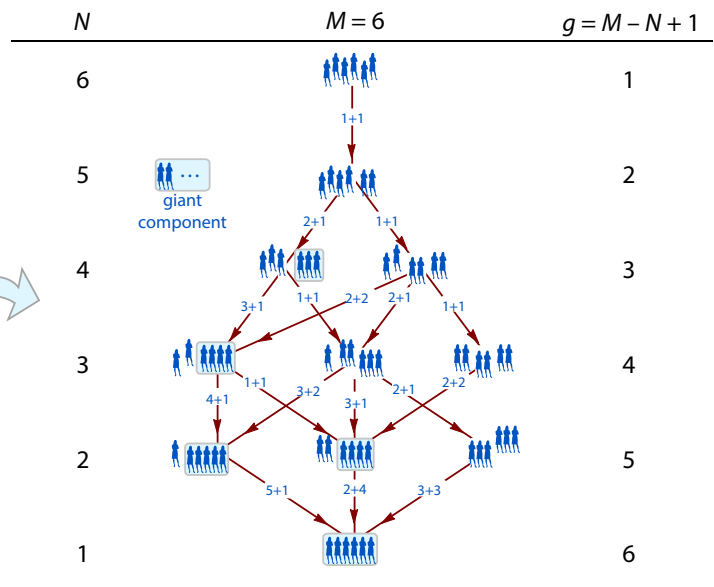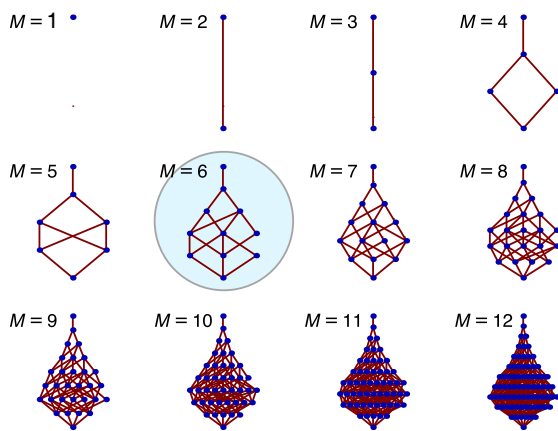

Figure 1 | The ensemble of discrete binary aggregation graphs ($M = 1$ through 12 shown). Each graph starts with M unattached monomers at the top and ends with a single cluster at the bottom. The graph $M = 6$ is shown in detail on the right. Nodes represent distributions; arrows point from the parent distribution to the offspring and are labelled by the sizes of the two clusters whose merging produces the offspring. All distributions at fixed N constitute the (M, N) ensemble. The giant component (highlighted) is identified as the cluster in the size region $i > (M - N + 1)/2 + 1$. At most one cluster can exist in the giant region[13].

$$\bar{k}_{\mathbf{n'}} = \frac{1}{N(N+1)} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} n'_i \big(n'_j - \delta_{ij}\big) k_{ij}. \qquad (10)$$

Combining (2), (9) and (7), we obtain the recursion

$$\frac{\Omega_{M,N+1}}{\Omega_{M,N}} = \frac{M-N}{N} \frac{1}{\langle k_{M,N+1}\rangle} \left\{ \frac{\langle k_{M,N+1}\rangle}{M-N} \sum_{i=2}^{\infty} n_i \sum_{j=1}^{i-1} \frac{k_{i-j,j}}{\bar{k}_{\mathbf{n'}}} \frac{W(\mathbf{n'})}{W(\mathbf{n})} \right\}, (11)$$

where $\langle k_{M,N+1}\rangle = \sum P(\mathbf{n'})\bar{k}_{\mathbf{n'}}$ is the ensemble average kernel over all distributions of the parent ensemble. Equation (11) applies to all distributions $\mathbf{n}$ of the (M, N) ensemble, and since the left-hand side is strictly a function of M and N, the same must be true for the right-hand side: the quantity in braces must be independent of $\mathbf{n}$. We further require homogeneous behavior in the thermodynamic limit, such that $\log \tilde{w}_i$ is an intensive function of M/N. This condition fixes the quantity in braces to be 1 and breaks equation (11) into two separate recursions, one for Ω and one for $W(\mathbf{n})$. The first recursion is (see Supplementary Information for details)

$$\frac{\Omega_{M,N+1}}{\Omega_{M,N}} = \frac{M-N}{N} \frac{1}{\langle K_{M,N+1}\rangle} \qquad (12)$$

and is readily inverted to produce

$$\Omega_{M,N} = \binom{M-1}{N-1} \prod_{L=N+1}^{M} \langle K_{M,L}\rangle. \qquad (13)$$

Thus we have the partition function in terms of M, N, and the product of all $\langle K\rangle$ from generation 1 up to the parent of the current generation. We note that the binomial factor is the total number of ordered partitions of integer M into N[24], also equal to the number of distributions in the (M, N) ensemble, each counted $\mathbf{n}!$ times. The second recursion gives the selection bias $W(\mathbf{n})$ in terms of the bias of all parents n' of distribution $\mathbf{n}$ (see Supplementary Information for details):

$$W(\mathbf{n}) = \frac{\langle k_{M,N+1}\rangle}{M-N} \sum_{i=2}^{\infty} n_i \sum_{j=1}^{i-1} \frac{k_{i-j,j}}{\bar{k}_{\mathbf{n'}}} W(\mathbf{n'}). \qquad (14)$$

Starting with $W = 1$ in generation 1 we may obtain, in principle, the bias of any distribution in the phase space. Returning to equation (12), we recognize the right-hand side as q, which produces the path equation of the process:

$$q = \frac{M-N}{N} \frac{1}{\langle k_{M,N+1}\rangle}. \qquad (15)$$

Equations (13) and (14), along with (3)–(6) constitute a closed set of equations for the MPD in the thermodynamic limit.

Product Kernel
We now apply the theory to obtain the solution to the product kernel $k_{ij} = ij$. In the thermodynamic limit, $\langle k_{M,N}\rangle \to \bar{k}_{\mathbf{n}} \to (M/N)^2$. With this result and Eqs. (4)–(5) we obtain the parameters of the sol (see Supplementary Information for details):

$$\beta = 2\theta - \log\theta, \qquad (16)$$

$$q = \theta(1-\theta), \qquad (17)$$

$$\tilde{w}_i = 2 \frac{(2i)^{i-2}}{i!}, \qquad (18)$$

with $\theta = 1 - N/M$. The MPD follows from equation (3):

$$\frac{\tilde{n}_i}{N} = \frac{2\theta}{1-\theta} \frac{(2\theta k)^{k-2}}{k!} e^{-2\theta k}. \qquad (19)$$

The MPD of the sol satisfies the extremum condition $d \log \Omega_{M,N} = 0$, but for the state to be stable we must also have $d^2 \log \Omega_{M,N} \leq 0$, or $(\partial \log q/\partial N)_M \leq 0$. Applying this stability condition to equation (17) we conclude that the range of stability is $0 \leq \theta < 1/2$ and that phase splitting must occur at $N^* = M/2$. This is the same as the gel point in the Smoluchowski equation with monodisperse conditions. We now proceed to obtain solutions in the post-gel region. Consider a two-phase system that contains mass M_{sol} in the sol, and $M_{\mathrm{gel}} = M - M_{\mathrm{sol}}$ in the gel ($N_{\mathrm{gel}} = 1$, $N_{\mathrm{sol}} = N - 1$[13]). As an equilibrium phase, the sol maximizes $\Omega_{M_{\mathrm{sol}},N-1} \approx \Omega_{M_{\mathrm{sol}},N}$. Its distribution, therefore, is given by equation (3) with θ replaced by $\theta_{\mathrm{sol}} = 1 - N/M_{\mathrm{sol}}$. To determine M_{sol} we recall that equation (17) must be satisfied, at all times. Since stability requires $N \leq M/2$, we must have $M_{\mathrm{sol}} = MN/(M - N)$. Finally, the gel fraction is $\phi_{\mathrm{gel}} = (M - M_{\mathrm{sol}})/M$, or

$$\phi_{\mathrm{gel}} = 2 - 1/\theta, \quad (\theta \geq 1/2). \qquad (20)$$

Thus we have the complete solution: in the pre-gel region ($\theta \leq 1/2$) the sol is given by equation (19); in the post-gel region ($\theta \geq 1/2$) it given by the same equation with θ replaced by $\theta_{\mathrm{sol}} = 1 - \theta$, and the gel fraction is obtained from equation (20).

We illustrate the theory with a numerical calculation for $M = 40$. This value is sufficiently small that we may enumerate all distributions on the aggregation graph and perform an exact calculation of the entire ensemble, yet large enough that the thermodynamic limit is approached to satisfactory degree (the phase space contains 37338 distributions). As a further test we conduct Monte Carlo (MC) simulations by the constant-volume algorithm described in Ref. 25. The simulations sample the vicinity of the MPD (not the MPD itself) from which the mean distribution is calculated. The exact calculation is done on the entire graph as follows. Starting with $W = 1$ in generation $g = 1$, we apply equation (14) to obtain the bias of all distributions in the next generation until the entire graph is computed. Next we calculate the partition function in each generation from the normalization condition $\Omega = \sum \mathbf{n}! W(\mathbf{n})$, and the probability of distribution from equation (2). With all probabilities known, the mean distribution and the ensemble average kernel are readily calculated, and the MPD is identified by locating the maximum $P(\mathbf{n})$. As a check, we calculate the partition function from equation (13) and confirm that for pre-gel states it agrees with the result from the normalization condition.

These calculations are compared in Figure 2, which shows selected distributions ranging from $N = 33$ (early stage of mostly small clusters) to $N = 6$ (nearly fully gelled). Since the MPD is an actual member of the ensemble, it contains integer numbers of clusters. The mean distribution is a composite of the entire ensemble and is not restricted to integer values. The giant cluster forms at $N^* = 22$ and its presence is seen very clearly in the MPD. The gel phase is less prominent in the mean distribution because its peak is smeared by lateral fluctuations. Not all distributions in the vicinity of the MPD contain a giant cluster; as a result, the gel fraction grows smoothy at the gel point. In the sol region $1 \leq (M - N + 1)/2$, the theoretical distribution from equation (19) and the mean distribution are in excellent agreement. The analytic result eventually breaks down when $N \to 1$ (the thermodynamic limit is violated at this point), yet even with N as small as 6, agreement with theory remains acceptable. The mean distribution from MC is practically indistinguishable from that by the exact calculation. This confirms the validity of equation (14), which forms the basis of the exact calculation.

Discussion
Our results make contact with several studies in the literature. A recursion for the partition function that is similar to that in equation (12) (different by a factor that is inconsequential for the statistics of the ensemble but crucial for thermodynamics to work) was obtained by Spouge[18,26,27] by a combinatorial derivation for kernels of the form
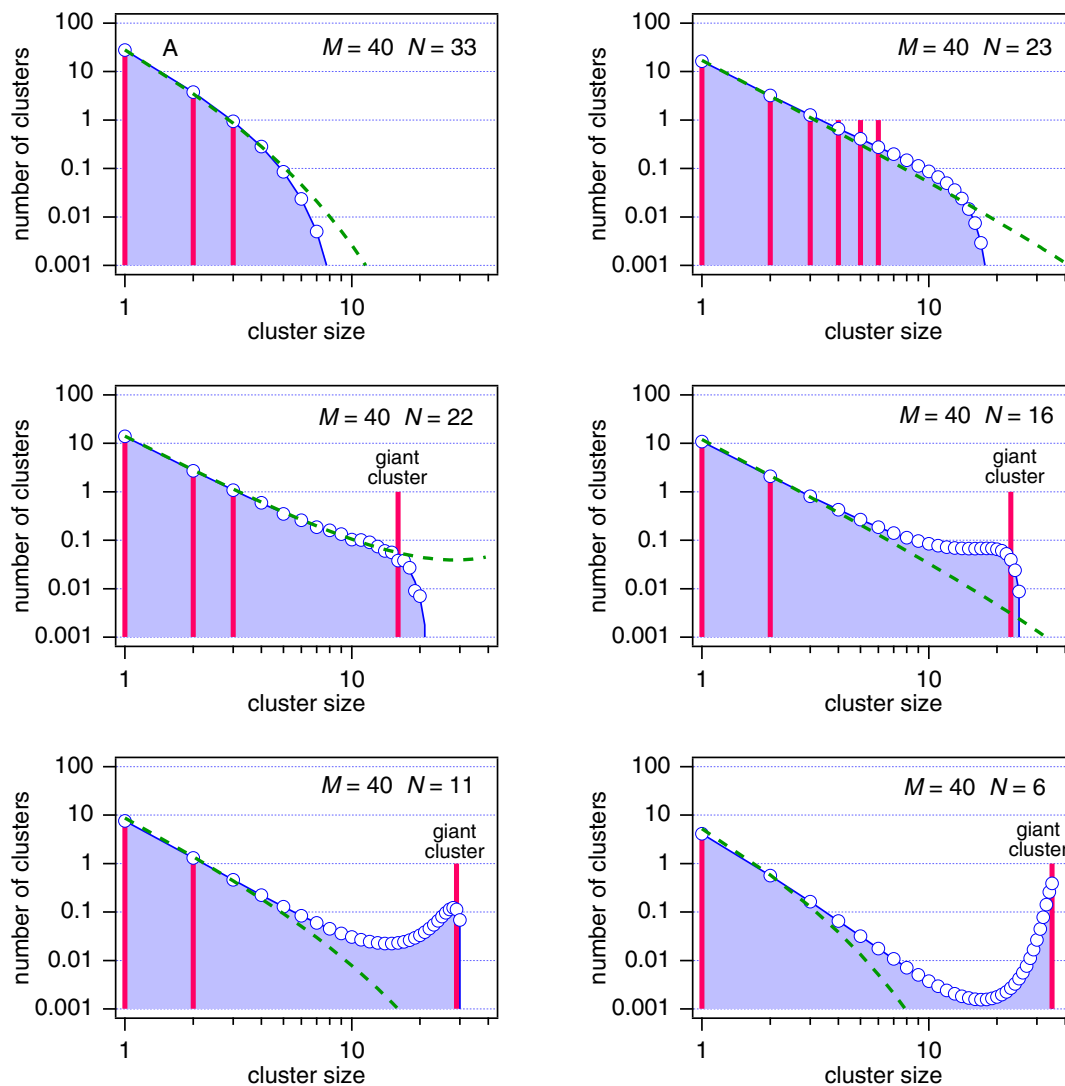
Figure 2 | Cluster distributions in a population with $M = 40$ as a function of the number of clusters N. Shaded curve: mean distribution (exact calculation by direct enumeration of all distributions); vertical sticks: most probable distribution (exact calculation); symbols: Monte Carlo simulation (average of 5000 repetitions); dashed line: equation (19) (thermodynamic limit) with $\theta = 1 - N/M$ in the pre-gel region ($N \geq M/2$), and $\theta = N/M$ in the post-gel region.

$k_{ij} = a + b(i + j) + c(ij)$ in pre-gel states. We recognize equation (19) as the classical pre-gel solution to the Smoluchowski equation[1,3]. We further recognize the post-gel solution as the Flory model, which assumes that the sol fraction continues to interact with the giant component past the gel point[12]. No such *a priori* assumption is required here: as long as no cluster is excluded from merging, a condition already built into the kernel ($k_{ij} \neq 0$ for all $i, j \geq 1$), the post-gel solution *is* the Flory solution.

We close with a final observation that points to an even closer analogy to molecular systems. Using equation (15) to calculate q we find

$$q = \left(\frac{M-N}{N}\right)\left(\frac{N+1}{M}\right)^2, \tag{21}$$

whose limiting value for $N \gg 1$ is the result given in equation (17). Plotted against $\theta = 1 - N/M$ over the full range $\theta = 0$ to 1, this equation shows behavior reminiscent of subcritical van der Waals isotherms (Fig. 3) and for large M it converges to a parabola in the region $0 \leq \theta < 1$, plus a Dirac delta function at $\theta = 1$. Stability requires $(\partial q/\partial \theta)_M \geq 0$, a condition that is met in $0 \leq \theta \leq 1/2$, but also on the Dirac branch. When the system crosses into the unstable

region (state A in Fig. 3) it must split into two phases. The sol phase is determined by equation (15), which produces a state on the stable branch ($\theta < 1/2$) at the same q. Extending this line to the right we
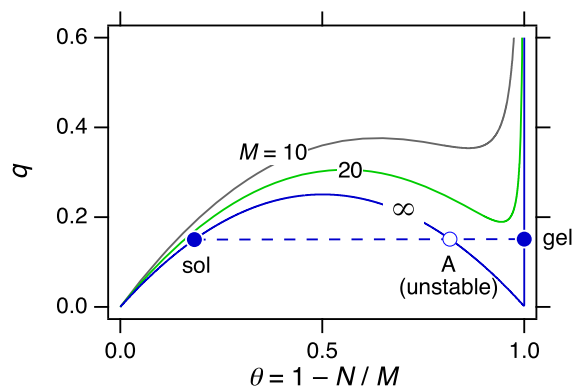

Figure 3 | Graphical constructions of tie line in the sol-gel region. State A is unstable and splits into a sol and a gel phase. All three states lie on the same tie line.

obtain an intersection with the Dirac delta branch, which we identify as the equilibrium gel phase at $\theta_{gel} = 1$. Thus we have the tie line of this two-phase system: it connects two equilibrium phases, with an unstable state at the middle.

The ensemble method was applied here to binary aggregation but can be adapted to any other growth mechanism. For example, by reversing the arrows in Fig. 1 we obtain the graph of binary fragmentation; by including both directions we obtain the graph for reversible aggregation/fragmentation (both processes share the same trajectories in phase space as binary aggregation). In general, the evolution of populations may be viewed as a swarm of trajectories in the phase space of Fig. (1) under parent-offspring relationships that must be derived for each case separately. We may draw, therefore, a rigorous connection between statistical thermodynamics and population balances that offers new insights into the dynamics of evolving populations.

1. Leyvraz, F. Scaling theory and exactly solved models in the kinetics of irreversible aggregation. *Phys. Rep.* **383**, 95–212 (2003).
2. Stockmayer, W. H. Theory of molecular size distribution and gel formation in branched-chain polymers. *J. Chem. Phys.* **11**, 45–55 (1943).
3. Aldous, D. J. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli* **5**, 3–48 (1999).
4. Smoluchowski, M. Versuch einer mathematischen theorie der Koagulationkinetic kolloider Loesungen. *Z. Phys. Chem.* **92**, 129–168 (1917).
5. Smoluchowski, M. Zusammenfassende Bearbeitungen. *Physik. Zeitschr.* **17**, 585 (1916).
6. Friedlander, S. K. *Smoke, Dust and Haze* (Oxford University Press, New York, 2000), second edn.
7. Gueron, S. & Levin, S. A. The dynamics of group formation. *Math. Biosci.* **128**, 243–264 (1995).
8. Drake, R. L. A mathematical survey of the coagulation equation. In Hidy, C. M. & Brock, J. R. (eds.) *Topics in current aerosol research*, vol. 3, 204–379 (Pergamon Press, New York, 1972).
9. Flory, P. J. Molecular size distribution in three dimensional polymers. II. trifunctional branching units. *J. Am. Chem. Soc.* **63**, 3091–3096 (1941).
10. Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17–61 (1960).
11. Krapivsky, P. L., Redner, S. & Ben-Naim, E. *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, 2010).
12. Ziff, R. M. & Stell, G. Kinetics of polymer gelation. *J. Chem. Phys.* **73**, 3492–3499 (1980).
13. Matsoukas, T. Statistical thermodynamics of clustered populations. *Phys. Rev. E* **90**, 022113 (2014).
14. Marcus, A. Stochastic coalescence. *Technometrics* **10**, 133–143 (1968).
15. Lushnikov, A. A. Coagulation in finite systems. *J. Colloid Interface Sci.* **65**, 276–285 (1978).
16. Lushnikov, A. A. From sol to gel exactly. *Phys. Rev. Lett.* **93**, 198302 (2004).
17. Lushnikov, A. A. Exact kinetics of the sol-gel transition. *Phys. Rev. E* **71**, 046129 (2005).
18. Spouge, J. L. Asymmetric bonding of identical units: A general $a_g r b_{f-g}$ polymer model. *Macromolecules* **16**, 831–838 (1983).
19. Hendriks, E. M., Spouge, J. L., Eibl, M. & Schreckenberg, M. Exact solutions for random coagulation processes. *Zeitschrift für Physik B Condensed Matter* **58**, 219–227 (1985).
20. Durrett, R., Granovsky, B. & Gueron, S. The equilibrium behavior of reversible coagulation-fragmentation processes. *J. Theor. Probab.* **12**, 447–474 (1999).
21. Berestycki, N. & Pitman, J. Gibbs distributions for random partitions generated by a fragmentation process. *J. Stat. Phys.* **127**, 381–418 (2007).
22. Duerr, H.-P. & Dietz, K. Stochastic models for aggregation processes. *Mathematical Biosciences* **165**, 135–145 (2000).
23. Sheth, R. K. & Pitman, J. Coagulation and branching process models of gravitational clustering. *Mon. Not. R. Astron. Soc.* **289**, 66–82 (1997).
24. Pitman, J. *Combinatorial Stochastic Processes*, vol. 1875 (Springer, 2006).
25. Smith, M. & Matsoukas, T. Constant-number Monte Carlo simulation of population balances. *Chem. Eng. Sci.* **53**, 1777–1786 (1998).
26. Spouge, J. L. Analytic solutions to Smoluchowski's coagulation equation: a combinatorial interpretation. *J. Phys. A* **18**, 3063 (1985).
27. Spouge, J. L. Equilibrium polymer size distributions. *Macromolecules* **16**, 121–127 (1983).

Additional information