Exactly solved models for irreversible aggregation

Seung-Woo Son,[1] Golnoosh Bizhani,[1] Peter Grassberger,[1,2] and Maya Paczuski[1]

[1] *Complexity Science Group, University of Calgary, Calgary T2N 1N4, Canada*
[2] *FZ Jülich, D-52425 Jülich, Germany*

(Dated: November 4, 2010)

We present the complete solution for the probability to find any given state in an aggregation process $(k+1)X \to X$, given a fixed number of unit mass particles in the initial state. For constant reaction rate, exactly the same probability is found in three cases: well-mixed solutions, particles on a ring reacting with k nearest neighbors to the right, and the same for particles on a line. The mass distribution of a single cluster exhibits scaling laws both at small and large sizes. We relate our findings to a fragmentation process via a nonlinear recursion relation as well as a type of percolation transition seen in network renormalization.

PACS numbers: 89.75.Hc, 02.10.Ox, 05.70.Ln

Droplets beget rain, goblets coagulate to make butter or cream, and dust particles stick together to form aggregates that can eventually coalesce into planets. At the microscopic level, irreversible aggregation of atoms and molecules creates many familiar forms of matter such as aerosols, colloids, gels, suspensions, clusters and solids [1]. Some catastrophic diseases, such ashemophilia [2] or neurodegeneration [3] are caused by defects in coagulation. But models of aggregation also turn up in abstract circumstances such as the formation of political parties [4], algorithms for decision making [5], data analysis [6], database construction [7], or random partitions applied to genetics [8–10] – to name a few.

Almost a century ago, Smoluchowski proposed a theory based on rate equations to describe processes governed by diffusion, collision and irreversible merging of aggregates [11]. The theory predicts how many small and large clusters exist at any given time. The mass distribution depends on certain details such as the initial conditions, reactions present, relative rates of aggregation compared to other processes, various mass dependencies of the rates, and the presence or absence of spatial structure. In general one observes in reaction-diffusion systems (such as e.g. DLA [12]) that a finite speed of diffusion typically leads to negative spatial correlations which in turn affect reaction dynamics. The Smoluchowski approach has been extended to include, for instance, fluctuations via a time dependent reaction kernel [13]. A key interest to physicists has been the derivation of scaling laws that characterize different universality classes [14].

Here we study stylized models of aggregation that throw out almost all details to focus exclusively on coalescence itself. We derive complete, exact solutions for the probability to find any given state in three models and show that these solutions are precisely the same. Hence the behavior is markedly robust – which is consistent with the generality of aggregation processes spanning both physical and non-physical domains. We further show how aggregation (when the parameter in the model k is even) is related to random sequential renormalization (RSR) of graphs [15] and how the scaling behavior observed reflects the presence of a novel percolation transition rather than an underlying fractal. The latter is discussed in the context of claims [16, 17] that complex networks exhibit non-trivial fractal dimensions.

We consider models governed by the reaction $(k+1)X \to X$, where a randomly picked cluster coalesces with k other ones. The mass of the newly formed cluster is the sum of the $(k+1)$ masses and the reaction rate is constant – independent of time, mass or any other quantity. We consider two extreme cases: well-mixed systems and one-dimensional models – both a ring with periodic boundary conditions and a line with open ones. In any case, reactions are allowed only if there is a sufficient number, k, of available clusters.

Let us start with the model defined on a ring. Initially, N_0 sites labelled by $i \in [1, ...N_0]$ are each occupied by a particle of mass $m = 1$. Time can be either discrete or

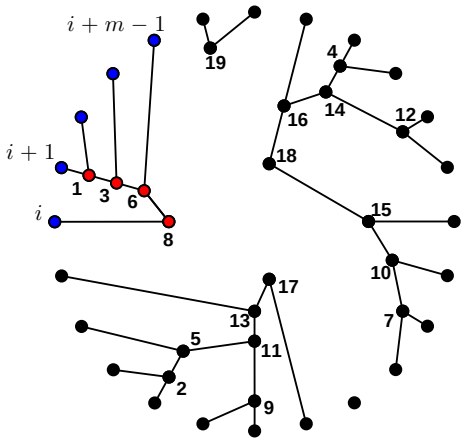

FIG. 1: (Color online) Illustration of aggregation on a ring with $k = 1$, $N_0 = 24$, and $N = 5$. The tree in color corresponds to a cluster of mass $m = 5$. It has five leaves (blue) and four internal nodes (red). Its leaves start at site i and end at site $i + m - 1$. The numbers beside internal nodes correspond to the time when coalescence occurs.

continuous, but we demand that two events never happen simultaneously. Hence, events can be ranked by increasing time, and are denoted by positive, integer values t. Due to these events, particles coagulate to form clusters of mass $m > 1$. More precisely, each event consists of picking a random cluster with uniform probability and joining it with k clusters to its immediate right, using periodic boundary conditions. For k even, the same results are found if we aggregate clusters symmetrically. After t events, $N_t = N_0 - kt$ clusters exist. Our main result is the probability to find any connected sequence of cluster masses $p_{N_t}^{N_0}(m_1, m_2 \dots m_{N_t})$ – where a cluster of mass m_1 is followed by a cluster of mass m_2, etc., moving clockwise (see Fig. 1). First we start with the single cluster mass probability.

Cluster masses are restricted to $m \equiv 1 \pmod k$. Defining $m - 1 = ks$, the integer s is the number of events needed to make the cluster of mass m. As depicted in Fig. 1, we can represent any realization of the process by a forest of N_t rooted trees with N_0 leaves and t internal nodes. Each tree α has s_α internal nodes, with $\sum_\alpha s_\alpha = t$. We simplify notation by writing N for N_t.

Let $\pi_N^{N_0}(m)$ denote for fixed k (the dependence on k is not written explicitly in the following) the probability that a cluster of mass m has its left-most member at site $i \in [1, N_0]$ after t events. The probability that any of the N clusters picked at random has mass m is then

$$p_N^{N_0}(m) = \frac{N_0}{N}\, \pi_N^{N_0}(m) \quad , \tag{1}$$

because there are N_0 choices for i and the chance to pick that particular cluster given that it exists is $1/N$. Since events occur completely at random, each *history* occurs with equal probability. The term 'history' refers to a fixed forest, which includes a fixed temporal order of events. Thus $\pi_N^{N_0}(m)$ is equal to the number of histories leading to a final configuration with a cluster of mass m starting at position i, divided by all possible histories leading to N clusters. The latter is equal to

$$n_{\text{hist,tot}} = N_0 \times (N_0 - k) \times \dots (N + k), \tag{2}$$

where each of the t factors equals the number of choices for the next event. Using Pochhammer $k-$symbols or, equivalently, generalized rising factorials [8, 9, 18, 19], this can be written as

$$n_{\text{hist,tot}} = (N + k)_{t,k} \quad . \tag{3}$$

Similarly, the number of histories leading to a cluster of size m starting at a fixed position i is

$$n_{\text{hist,cluster}} = (m - k)(m - 2k) \times \dots 1 = (1)_{s,k} \tag{4}$$

and the number of histories for the remaining $N-1$ clusters is

$$\begin{aligned} n_{\text{hist,rest}} &= (N_0 - m - k)(N_0 - m - 2k) \times \dots (N - 1) \\ &= (N - 1)_{t-s,k} \quad . \end{aligned} \tag{5}$$

So far we have not included the number of choices associated with different time orderings for the s events in the cluster and $(t - s)$ events in the rest of the forest. The number of different time orderings is given by

$$n_{\text{orderings}} = \binom{t}{s} \quad . \tag{6}$$

Combining Eqs. (1) to (6), we obtain

$$\begin{aligned} p_N^{N_0}(m) &= \frac{N_0}{N}\binom{t}{s}\frac{(N-1)_{t-s,k}(1)_{s,k}}{(N+k)_{t,k}} \\ &= \binom{t}{s}\frac{(N-1)_{t-s,k}(1)_{s,k}}{(N)_{t,k}}. \end{aligned} \tag{7}$$

This result can be further simplified into beta functions or, more conveniently, k-beta functions (see e.g. [19]),

$$B_k(x, y) = \frac{1}{k} B(\frac{x}{k}, \frac{y}{k}) \quad ,$$

giving a remarkably simple final result

$$p_N^{N_0}(m) = \binom{t}{s}\frac{B_k(N_0 - m, m)}{B_k(N - 1, 1)} \quad . \tag{8}$$

We make a number of observations: (1) For $k = 1$ the process maps to bond percolation on a ring. For $N = 2$, the mass distribution is uniform over the entire range $m \in [1, N_0 - 1]$. For $N > 2$, the distribution is proportional to the $(N - 2)^{nd}$ factorial power $((N_0 - m - 1)(N_0 - m - 2) \cdots (N_0 - m - N + 2))$. (2) For $N = 2$ and any $k \geq 1$, $p_N^{N_0}(m)$ is symmetric under the exchange $m \leftrightarrow N_0 - m$. (3) For $N = 2$ and $k = 2$ we obtain an equation formally identical to Spitzer's discrete arcsine law for fluctuations of random walks [20]. (4) Asymptotic power laws for $N_0 \to \infty$ can be determined using Stirling's formula. If N is fixed and both m and $(N_0 - m) \to \infty$,

$$p_N^{N_0}(m) \sim \frac{(t - s)^{\frac{N-1}{k}-1}}{s^{1-\frac{1}{k}}} \quad . \tag{9}$$

For small masses, this gives a decreasing power law with exponent $-1 + 1/k$. For $N = k + 1$, the power law $p_N^{N_0}(m) \sim s^{-1+1/k}$ holds up to the largest possible value, $m = N_0 - N + 1$, and the cutoff is a step function. For $m/N_0 \to 1$ different power laws appear if $N \neq k + 1$, and the sign of the exponent changes at $N = k + 1$. For $N < k + 1$, the distribution has a peak at $m/N_0 \to 1$, while it goes to zero for $N > k + 1$. These scaling laws are illustrated for $k = 2$ in Fig. 2. (5) The scaling laws

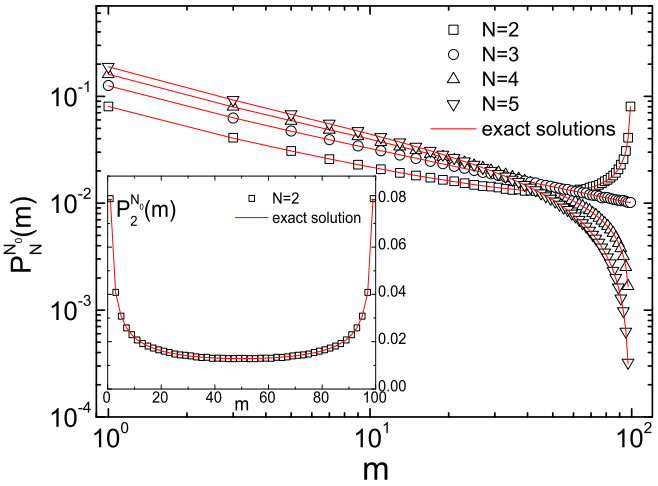

FIG. 2: (Color online) Cluster size distributions after $t = 50$ events for $k = 2$, for different values of N averaged over 10^6 realizations compared to exact results. The large size behavior changes from increasing power law to a decreasing one at $N = k + 1$. The inset shows the symmetric discrete arcsine law found for $N = 2$.

found for $m \ll N_0$ are identical to those obtained by Krapivsky [21] for the well mixed case. However, the behavior for $m/N_0 \to 1$ given in [21] does not agree with our result. (6) The probability $p_N^{N_0}(m)$ satisfies a number of recursion relations:

$$p_N^{N_0}(m+k) = \frac{m(N_0 - m - N + 1)}{(m+k-1)(N_0 - m - k)}\, p_N^{N_0}(m)\ ,$$

$$p_{N+k}^{N_0}(m) = \frac{N(N_0 - m - N + 1)}{(N-1)(N_0 - N)}\, p_N^{N_0}(m)\ . \qquad (10)$$

A third *nonlinear* recursion relation is given later.

Joint distributions for masses of adjacent clusters can also be found. We denote by $p_N^{N_0}(m_1, m_2)$ the probability to find a cluster of mass m_1 followed immediately to the right by a cluster of mass m_2. This is non-zero only if $m_1 = 1 + s_1 k$ and $m_2 = 1 + s_2 k$, where s_α is the number of events needed to form a cluster of mass m_α. By the same arguments that led to Eq. (7) we get

$$p_N^{N_0}(m_1, m_2) = \frac{\mathcal{T}[t, \{s\}, 3](N-2)_{s_0,k}(1)_{s_1,k}(1)_{s_2,k}}{(N)_{t,k}}\ ,$$

where $s_0 = t - \sum_{\beta=1}^{\alpha} s_\beta$ and the multinomial coefficient

$$\mathcal{T}[t, \{s\}, \alpha+1] = \begin{pmatrix} t \\ s_0, \ldots s_\alpha \end{pmatrix}.$$

When $\alpha = 2$, it is a trinomial coefficient that counts the number of ways in which the three sequences of events – for the two clusters considered, and for all $(N-2)$ other clusters – can be interleaved in a single history.

For any $1 \le \alpha \le N-1$ the joint probability distribution for α consecutive, adjacent clusters is a product of a multinomial coefficient and $\alpha + 1$ Pochhammer k-symbols, divided by the Pochhammer k-symbol related to the total number of possible histories given N_0 initial particles. Defining again s_0 as the number of events in all clusters except the first α ones, we can write the result compactly as

$$p_N^{N_0}(m_1, \ldots m_\alpha) = \frac{\mathcal{T}[t, \{s\}, \alpha+1](N-\alpha)_{s_0,k} \prod_{\beta=1}^{\alpha}(1)_{s_\beta,k}}{(N)_{t,k}}. \qquad (11)$$

In particular, this can also be done for the joint distribution for all N masses by setting $\alpha = N - 1$. The resulting expression is then manifestly invariant under any permutations of N numbers $(m_1, \ldots m_N)$. Hence the $N-$ cluster probability is independent of the spatial ordering of the clusters. While there are obvious correlations between the mass values (the sum of all cluster masses must be N_0), there are *no spatial correlations*.

We now consider a line of N_0 particles with open boundaries. Again, aggregation events consist of a random choice of a cluster, followed by its amalgamation with its k nearest neighbors to the right. The target cluster must be at least k steps away from the right-most boundary. Following the same arguments leads immediately to Eq. (11) for $\alpha = N - 1$, showing that the two models lead to precisely the same statistics.

The absence of spatial correlations indicates that the same dynamics might result for the well-mixed case. Now we start with a bucket containing N_0 balls, each of unit mass. An event consists of taking $k + 1$ balls out of the bucket, merging them together, and returning the new ball to the bucket. The $k+1$ balls are chosen completely at random, independent of their masses.

The single cluster mass distribution for the well-mixed model can be obtained using the same strategy as before, but the details are quite different. Consider the total number of histories. Since events now correspond to choosing any $k+1$ balls out of $N_0 - kt$ balls, we have, instead of the Pochhammer k-symbol, a product of binomial coefficients,

$$n_{\rm hist,tot} = \binom{N_0}{k+1}\binom{N_0-k}{k+1}\cdots\binom{N+k}{k+1}. \qquad (12)$$

The expressions for $n_{\rm hist,cluster}$ and $n_{\rm hist,rest}$ are analogous, with the factors $(m - jk)$ (resp. $(N_0 - m - jk)$) in Eq. (4) (resp. (5)) replaced by binomial coefficients. The number of time orderings $n_{\rm orderings}$ is exactly the same as before, but the first factor N_0/N in Eq. (7) has to be replaced by $\frac{1}{N}\binom{N_0}{m}$. Putting all these things together, many cancellations take place, leading exactly to Eq. (8). This argument can be similarly extending to get the full N-particle distribution function, obtaining exactly the same result as before, for any k.

The time reversed process of aggregation is fragmenta-

tion. When considering the fragmentation process associated with any of these models, we have to carefully evaluate fragmentation rates. Assuming uniform rates would not lead to all time reversed histories having the same probability. Indeed the fraction of all mergers associated with making a cluster of mass m' is $(s'/t) = (m'-1)/(N_0-N)$, which must equal the probability that an existing cluster of mass m' will fragment at the next step in the time reversed process. If it does, then for consistency its fragmentation products must have a mass distribution given by $p_{k+1}^{m'}(m)$. A quadratic recursion relation for $p_{N+k}^{N_0}(m)$ can then be obtained by considering the likelihood of all fragmentation events in a configuration of N clusters, with m being the mass of one of the resulting $k+1$ fragmentation products. The relation is

$$p_{N+k}^{N_0}(m) = \sum_{m'=m+k}^{N_0-N+1}{}' \frac{N(m'-1)p_N^{N_0}(m')p_{k+1}^{m'}(m)}{N_0-N},$$

where the prime on the summation symbol indicates that m' must increase in steps of k.

The present work started as part of an ongoing investigation of renormalization schemes for complex networks. This was triggered by claims that one can define finite fractal dimensions for real world networks such as the world-wide-web [16] – a surprising assertion, are given the fact that these networks typically small world [22, 23]. In particular, the number of nodes increases roughly exponentially (or even faster than exponentially [24, 25]) with distance from the target node. The renormalization group (RG) flow underlying the box covering of [16, 17] was studied carefully in [26, 27], where it was found that the scaling laws may be related to an "RG fixed point" which was observed for a wide variety of networks. Ref. [15] introduced a fully sequential scheme (called random sequential renormalization, RSR) which eliminates formidable difficulties associated with covering a network with boxes.

Results for various graphs (critical trees [15], Erdös-Renyi and Barabasi-Albert [28] networks [29], regular lattices [30]) indicate that a fixed point, associated with scaling behavior, exists. It is a novel type of percolation transition (in general in a new universality class) that it is not contingent upon fractality of the original graph. The aggregation process on a one dimensional ring or line (with k even) studied here is precisely the RSR on these lattices. The non-trivial scaling laws we find for small and large cluster sizes indicates that also in this simple geometry a non-trivial fixed point exists that it is not in the ordinary percolation universality class (which would correspond to $k=1$) and that does not reflect any underlying fractality of the graph.

In summary, we have given complete solutions to find any state for a class of aggregation models starting with uniform initial conditions. The fact that we could solve exactly a one dimensional model without detailed balance might seem surprising since such models are in general not solvable, unless they satisfy detailed balance. It is related to the fact that spatial correlations, although they seem *a priori* not to be excluded, are in fact absent. Related to this is our finding that the well-mixed models have exactly the same solutions. It would be interesting to see whether generalizations of these observations hold true for more complicated models. For instance, the coagulation rates could depend on the cluster masses, in which case the sums over histories are weighted path integrals. We have also extended the application of aggregation models to renormalization schemes for complex networks.

[1] A. Zangwill, Nature **411**, 651 (2001).
[2] G. Born and M. Cross, J. Physiology **168**, 178 (1963).
[3] M. Bucciantini et al., Nature **416**, 507 (2002).
[4] R. Michels, *Political parties: A sociological study of the oligarchical tendencies of modern democracy* (Free Press, 1966).
[5] R. Yager, IEEE Trans. Systems, Man and Cybernetics **18**, 183 (1988).
[6] C. Intanagonwiwat, D. Estrin, R. Govindan, and J. Heidemann, in *Distributed Computing Systems, 2002. Proceedings. 22nd International Conference on* (IEEE, 2002), pp. 457–458.
[7] J. Smith and D. Smith, ACM TODS **2**, 105 (1977).
[8] J. Pitman and J. Picard, *Combinatorial stochastic processes* (Springer, 2006).
[9] J. Kingman, J. Appl. Prob. **19**, 27 (1982).
[10] M. Nordborg, in *Handbook of Statistical Genetics*, edited by D. J. Balding, M. J. Bishop, and C. Cannings (John Wiley & Sons, 2001), pp. 179–212.
[11] M. von Smoluchowski, Z. Phys. Chem. **92**, 129 (1917).
[12] P. Meakin, Phys. Rev. Lett. **51**, 1119 (1983).
[13] K. Kang and S. Redner, Phys. Rev. A **30**, 2833 (1984).
[14] F. Leyvraz, Phys. Rep. **383**, 95 (2003).
[15] G. Bizhani, V. Sood, M. Paczuski, and P. Grassberger, e-print arXiv:1009.3955, to appear in Phys. Rev. E (2010).
[16] C. Song, S. Havlin, and H. Makse, Nature **433**, 392 (2005).
[17] J. S. Kim, K. I. Goh, G. Salvi, E. Oh, B. Kahng, and D. Kim, Phys. Rev. E **75**, 16110 (2007).
[18] J. Normand, J. Phys. A **37**, 5737 (2004).
[19] R. Díaz and E. Pariguan, Divulg. Mat. **15**, 179 (2007).
[20] F. Spitzer, *Principles of random walk* (Springer Verlag, 2001).
[21] P. Krapivsky, J. Phys. A **24**, 4697 (1991).
[22] S. Milgram, Psychology Today **2**, 60 (1967).
[23] D. Watts and S. Strogatz, Nature **393**, 440 (1998).
[24] F. Chung and L. Lu, Internet Math. **1**, 15879 (2002).
[25] R. Cohen and S. Havlin, Phys. Rev. Lett. **90**, 58701 (2003).
[26] F. Radicchi, J. J. Ramasco, A. Barrat, and S. Fortunato, Phys. Rev. Lett. **101**, 148701 (2008).
[27] F. Radicchi, A. Barrat, S. Fortunato, and J. J. Ramasco,

Phys. Rev. E **79**, 26104 (2009).
[28] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[29] G. Bizhani et al., in preparation (2010).
[30] C. Christensen et al., in preparation (2010).