Relating the microscopic rules in coalescence-fragmentation models to the emergent cluster-size distribution

Błażej Ruszczycki[1], Ben Burnett[2], Zhenyuan Zhao[1], and Neil F. Johnson[1]

[1]*Physics Department, University of Miami, Coral Gables, FL 33146, USA*

[2]*Rudolf Peierls Centre for Theoretical Physics, Oxford University, Oxford, U.K.*

(Dated: January 30, 2009)

Coalescence-fragmentation problems are of great interest across the physical, biological, and recently social sciences. They are typically studied from the perspective of the rate equations, at the heart of such models are the rules used for coalescence and fragmentation. Here we discuss how changes in these microscopic rules affect the macroscopic cluster-size distribution which emerges from the solution to the rate equation. More generally, our work elucidates the crucial role that the fragmentation rule can play in such dynamical grouping models. We focus on two well-known models whose fragmentation rules lie at opposite extremes setting the models within the broader context of binary coalescence-fragmentation models. Further, we provide a range of generalizations and new analytic results for a well-known model of social group formation [V. M. Eguíluz and M. G. Zimmermann, Phys. Rev. Lett. **85**, 5659 (2000)]. We develop analytic perturbation treatment of the original model, and extend the mathematical to the treatment of growing and declining populations.

I. INTRODUCTION

The challenge to understand the dynamics of Complex Systems is attracting increasing attention, particularly in the socio-economic and biological domains [8, 11, 15, 16, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39]. For example, the recent turmoil in the financial markets has created significant public speculation as to the root cause of the observed fluctuations. At their heart, all Complex Systems share the common property of featuring many interacting objects from which the observed macroscopic dynamics emerge. Exactly how this happens cannot yet be specified in a generic way – however, an important milestone in this endeveour is to develop a quantitative understanding of any internal clustering dynamics within the population. Coalescence-fragmentation processes have been studied widely in conventional chemistry and physics [1, 2, 3, 4, 5, 6, 7, 9, 10, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52] – however, collective behavior in social systems is not limited by nearest neighbor interactions, nor are the details of social coalescence or fragmentation processes necessarily the same as in physical and biological systems. The challenge for a theorist is then twofold: (1) to provide a model which accounts correctly for the observed real-world behaviour — i.e., in the case that power-laws are observed empirically, the model should be able to reproduce the power-law dependence itself, the value of the corresponding power-law exponent, and possibly also the form of the truncation; (2) the rules invoked in the model need to make sense in the context of the real-world system being discussed.

In this paper, we discuss coalescence and fragmentation problems with a focus on social systems in which a typical fragmentation process corresponds to an entire cluster breaking up, as opposed to the typical binary splitting studied in physical and biological systems. In order to clarify the issues with the particular choice of the fragmentation we compare two well-known coalescence-fragmentation models, with fragmentation rules which lie at opposite extremes (Sec. IV). One of these is a well-known physics-inspired model of social group formation [V. M. Eguiluz and M. G. Zimmermann, Phys. Rev. Lett. 85, 5659 (2000)], while the other is a standard model in mathematical ecology due to Gueron and Levin [Gueron and Levin, Mathematical Biosciences 128, 243–246 (1995)]. This comparison between the two models allows us to elucidate the subtle differences in their microscopic rules that make their macroscopic dynamics differ, and leads us to a better generic understanding of the crucial role that the fragmentation rule can play. We then proceed to focus on the physics-based model of Eguiluz and Zimmermann, generalizing it in several ways and providing new analytic results (Sec. V), analyzing the perturbed version of the Eguiluz-Zimmerman model where the spontaneous cluster formation is present (Sec. V A), the models with steadily growing (Sec. V C 1) and declining population (Sec. V C 2). Further realistic modifications of the Eguiluz-Zimmerman model are discussed in Sec. V D.

There is of course a huge volume of work in the mathematics, physics and chemistry literature on the topic of clustering within a many-body population of interacting particles. Among these, the Smoluchowski coalescence equation is arguably the most famous and well-studied example [1, 2, 3]. Several interesting studies have tended to focus on important mathematical issues such as existence, uniqueness, mass conservation, gelation and finite size effects (see [4, 5, 6, 7, 9, 10] and references therein). When it comes to Complex Systems – and in particular, social systems – the more pressing goal is to understand the emergent dynamics of the population. In contrast

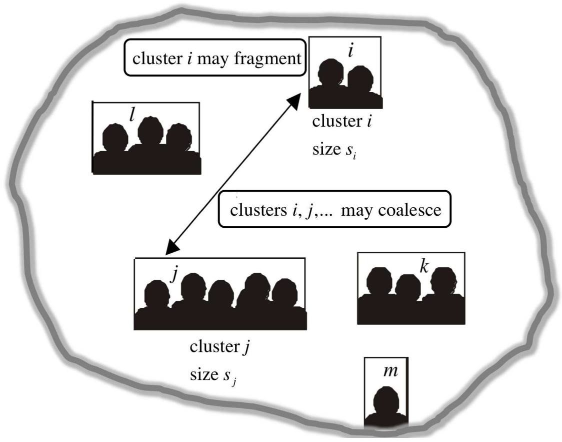

FIG. 1: Schematic diagram indicating the presence of coalescence and fragmentation processes, for a population of $N = 15$ objects dynamically partitioned into clusters. The size of cluster i is $s_i = 2$, while the size of cluster j is $s_j = 6$ etc. The fragmentation rule will generally exhibit the richest range of possibilities, given the combinatorial way in which a given cluster can fragment. There are many possible realizations of the objects themselves, e.g. humans, animals, macromolecules, though for simplicity we show them as humans.

to physical and chemical systems in which collision energetics play a crucial role in guiding the specification of microscopic coalescence and fragmentation rules, the precise microscopic rules in social systems are unknown – however, the overall macroscopic emergent phenomena such as cluster size distribution can be measured relatively easily. For example, in financial markets, the collective dynamics of the population of traders is registered directly by means of the price. Indeed, as many prior works have shown, such collective behavior in social systems tends to produce near scale-free (i.e. power-law) networks and/or cluster sizes in a variety of real-world situations. In financial markets, the volume of trades follows a power-law with slope near 2.5. Here we focus on the collective dynamics of such social systems by means of coalescence-fragmentation models which are appropriate to social systems – in so doing, we connect our discussion to the wider existing literature on coalescence-fragmentation models in physical and biological systems.

II. MODELING SOCIAL SYSTEMS

Many social systems can be viewed as comprising a large number of dynamically evolving clusters that can coalesce with other clusters to form even larger clusters, or fragment to form a collection of smaller ones. Figure 1 illustrates the situation of interest in many recent works on coalescence-fragmentation models[11, 16, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39]. As a result of coalescence and fragmentation processes over time, the population of N objects undergoes dynamical partitioning into clusters $i, j, k, \ldots$ of size $s_i, s_j, s_k, \ldots$, where both the number of clusters and their membership are typically time-dependent. We have denoted the N objects in human form, but of course they could be animals, macromolecules or other indivisible entities. Earlier studies tended to focus on situations in which the interactions between clusters might be expected to decay with physical separation – as in a simple solution of molecules interacting through Van der Waals interactions for example. However in social applications, where long-distance communication is as commonplace as communication with neighbors, it makes more sense to have interactions over all lengthscales, with the interaction probability effectively independent of physical separation.

Of the two processes in Fig. 1, i.e. coalescence and fragmentation, the coalescence process is likely to be the simpler and more generic. Suppose we have a particular partition of a population of N objects into clusters as in Fig. 1, and that a cluster i of size $s_i = 2$ is to coalesce. It is unlikely to undergo three-body collisions and/or interactions, and hence its most likely coalescence event is to join with a single other cluster j. Given that the size of a cluster measures the number of objects in it, it is therefore reasonable to imagine that the cluster probability should increase as the size of the clusters themselves increase. In a more human setting, the more objects that a cluster contains, the more likely it is that something will happen to one of its members in order to induce such an event.

Although we are using the term 'cluster' throughout this paper for convenience, it can also be taken to mean a 'community' in the language of network science[19] since it denotes a subset of the population who have very strong links between them, while the links between clusters are negligibly weak. Note also that the term 'cluster' need not necessarily mean physical connection – instead it could represent a group of objects whose actions happen to be coordinated in some way. Hence the coalescing of two clusters, however distant in real space, can mean an instantaneous alignment of their coordinated activities, as one might expect in a financial market[16], organized crime or insurgent warfare[15]. Likewise, fragmentation is then taken to mean instantaneous disruption of this coordination. Although we do not explore such specific applications here, it is useful to keep them in mind when we discuss the consequences of different fragmentation rules later in the paper.

The distinct feature of many real-world systems is the existence of scale-free behavior in the time-averaged cluster size distribution[11, 15, 16], such that in the first instance these systems can be characterized by the exponent of their power law and by the range of its scale-free behavior. One may therefore ask: Which ingredients of the coalescence-fragmentation models control the various observable aspects? In general, this question may not be well-specified since it might be a particular combination of factors, rather than a single ingredient, that is responsible for some particular behavior. It is these issues that we examine in this paper.

III. GENERAL FORMULATION

Once the probabilities specifying the coalescence and fragmentation are given, the cluster size distribution may be computed either by a direct simulation of the model or in a mean-field theory approximation by solving an appropriate set of rate equations, often numerically. The rate equations are typically non-linear and subsequently the non-trivial question of existence and uniqueness of the time-independent solution arises, these issues are addressed in foundational works as [12, 13, 14]. For several cases of interest of social/economical models the uniqueness and existence is be shown on the level of rate equations and identified via direct simulations. We will consider mostly "steady-state" models, in which there is some meaningful long-time behavior.

We denote by n_s the time-average number of clusters of size s and by N the total number of members (i.e. the population size). In order to characterize a general system, we need to prescribe the following two functions, each of dimension time^{-1}:

- $\mathcal{C}(s, s')$, the *coalescence function* which is the rate of merging of two clusters of sizes s and s' merge. We consider the coalescence which depends only on details pertaining exclusively to each pair of clusters therefore the possibility that 3 (or more) clusters are involved in the merging process is excluded.

- $\mathcal{FR}(s; m_1, m_2, \ldots, m_n - 1)$, the *fragmentation function* which is the rate describing the process by which a cluster of size s fragments into a configuration which contains m_1 clusters of size 1, m_2 clusters of size 2, etc.

The functional form of the above two functions is taken to be time-independent. If we consider general fragmentation processes, we see that a large number of parameters are necessary to characterize the fragmentation. However in order to write down the rate equations and hence calculate the cluster size distribution, we do not need complete knowledge of the fragmentation function (i.e. we do not need knowledge about all possible partitions). It is sufficient to know the *reduced fragmentation function* $\mathcal{F}(s, s', m)$, defined as the rate that a cluster of size s fragments into a configuration which contains m clusters of size s plus any other clusters of sizes different than s'. In addition to $\mathcal{F}(s, s', m)$ we need to know the rate that the fragmentation of any given cluster of size s occurs, which we denote as $f(s)$ – in principle we can calculate it by summing the complete fragmentation function over all partitions of the fragmentation products. We stress that by prescribing the deduced fragmentation function $\mathcal{F}(s, s', m)$ we do not characterize uniquely the fragmentation of the system and in general we may not be able to calculate $f(s)$ – yet it is possible in specific cases to do so once the assumption regarding the fragmentation products has been stated. If we look at the average number of clusters of size s that in unit time undergo the various processes (see Fig.2) we may introduce the following notation:

- $L_F(s)$: *loss due to fragmentation*, the number of clusters of size s that fragment
- $L_C(s)$: *loss due to coalescence*, the number of clusters of size s that join with other clusters
- $G_C(s)$: *gain from coalescence*, the number of clusters of size s created from the merging of clusters of size smaller than s
- $G_F(s)$: *gain from fragmentation*, the number of clusters of size s created from fragmenting clusters of size larger than s

Symbolically the rate equations for any s are written as

$$0 = -L_F(s) - L_C(s) + G_C(s) + G_F(s) \tag{1}$$

which explicitly reads as

$$0 = -f(s)\, n_s - n_s \sum_{s'=1}^{N} n_{s'} \mathcal{C}(s, s') + \frac{1}{2} \sum_{s'=1}^{s-1} n_{s'} n_{s-s'} \mathcal{C}(s', s-s') + \sum_{s+1}^{N} n_{s'} \sum_{m=1}^{[N/s']} m\, \mathcal{F}(s', s, m)\,. \tag{2}$$

The last term represents the gain in the number of clusters of size s coming from fragmentation of other clusters of size $s' > s$, in such a way that among the fragmentation products we have m clusters of size s. We are summing over all possible values of m and s'. Note that we sum over s' which is here the first (not the second) argument of $\mathcal{F}$. It is convenient to formally define

$$\widetilde{\mathcal{F}}(s, s') = \sum_{m=1}^{[N/s']} m\, \mathcal{F}(s, s', m)\,. \tag{3}$$

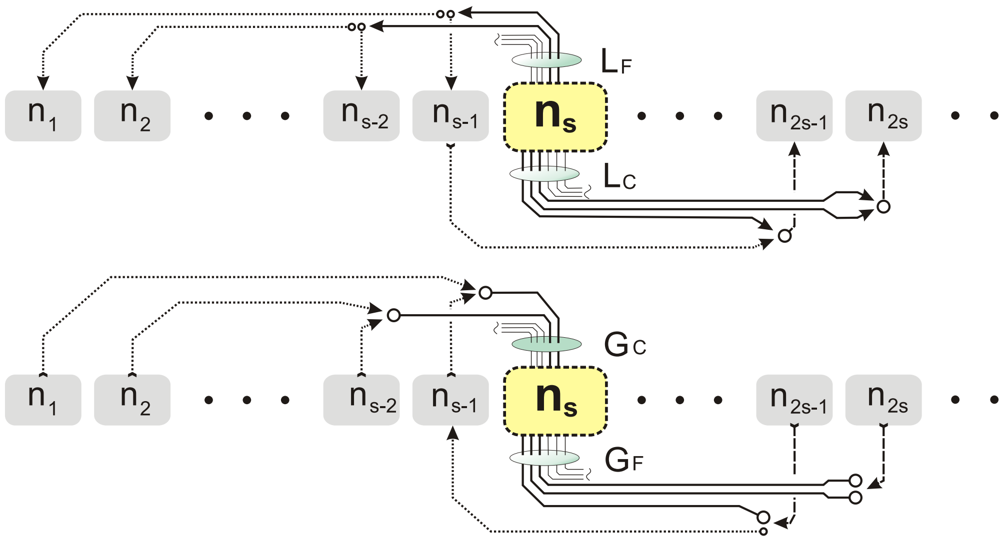

FIG. 2: The various processes of cluster coalescence and fragmentation which give rise to L_F, L_C, G_F, G_C for any particular value of s. The top figure represents the appearance of new clusters of size s, the bottom one represents their loss. In the interests of simplicity, the fragmentation into two clusters has been depicted and only a few processes are shown.

We write therefore the last term of Eq. (2) as

$$\sum_{s'=s+1}^{N} n_{s'}\widetilde{\mathcal{F}}(s', s) \, .$$

IV. ROLE OF THE FRAGMENTATION FUNCTION

A logical first step in the quest to understand classes of models which differ in their cluster fragmentation process, is to look at extreme cases. One such case is the Eguíluz-Zimmermann (E-Z) model [11]. In the E-Z model, fragmentation of a cluster of size s always produces s clusters of size 1, i.e. the cluster breaks up into individual objects. At the other extreme, is the famous Gueron and Levin (G-L) model [27] in which fragmentation of a cluster yields two smaller pieces, i.e. the original cluster splits into two clusters. The original G-L model is formulated in terms of continuous distributions – however, since our aim is to analyze the effects of these rules on the same footing, we will focus on the discrete version of the G-L model, returning to the continuous formulation later on. The common feature of the models that we discuss, is the presence of a separable coalescence function:

$$\mathcal{C}(s, s') = \alpha\, a(s)\, a(s') \quad . \tag{4}$$

In principle, the multiplicative constant may be absorbed into $a(s)$, however we prefer to keep it explicitly and adopt a dimensionless $a(s)$. This class of model is further specified by introducing a coalescence mechanism on the microscopic scale, namely that two clusters merge when any member from one cluster connects to any member from the other cluster. In a macroscopic description, this is equivalent to assuming that $a(s) = s$. We note that Gueron and Levin[27] having the solution of the rate equations for $a(s) = s$ considered explicitly other cases $a(s) = 1$ and $a(s) = 1/s$ by the substitution $n_s \to a(s)\, n_s$, this substitution however affects the form of the fragmentation function $\mathcal{F}(s, s', m)$

A. Fragmentation function

Assuming that the cluster may only split into two pieces still does not uniquely specify the fragmentation – we still need information about the probability distribution for the sizes of the fragments. In the G-L model, it is stated that the conditional distribution for fragments is uniform[27], i.e. the fragmentation of a cluster occurs with a probability which is independent of the way in which the cluster breaks. The reduced fragmentation function for $s > 1$ is therefore

$$\mathcal{F}_{GL}(s, s', m) = \beta\, b(s)\, [2\, \delta_{m,1}(1 - \delta_{2s',s}) + \delta_{m,2}\delta_{2s',s}] \tag{5}$$

where we have accounted for the fact that if $2s' = s$, the cluster breaks into two fragments of equal size. Using Eq. (3) one obtains immediately

$$\widetilde{\mathcal{F}}_{GL}(s, s') = 2\beta\, b(s) \quad . \tag{6}$$

The fragmentation probability is calculated as follows:

$$f_{GL}(s) = \frac{1}{2} \sum_{s'=1}^{s-1} \mathcal{F}_{GL}(s, s', m = 1) + \sum_{s'=1}^{s-1} \mathcal{F}_{GL}(s, s', m = 2) = \beta\, (s - 1)\, b(s)\,, \tag{7}$$

where the factor $1/2$ in the first term appears in order to avoid double-counting, and the second term represents splitting into two equal parts. In the E-Z fragmentation scheme, the cluster of size s can only break up into individual objects and there is only one mode of fragmentation, hence

$$\mathcal{F}_{EZ}(s, s', m) = \beta\, b(s)(1 - \delta_{s1})\delta_{s',1}\delta_{m,s}\,. \tag{8}$$

Using Eq. (3) we have

$$\widetilde{\mathcal{F}}_{EZ}(s, s') = \beta s\, b(s)\, (1 - \delta_{s1})\delta_{s',1}\,. \tag{9}$$

The fragmentation probability is

$$f_{EZ}(s) = \sum_{s'=1}^{s-1} \mathcal{F}_{EZ}(s, s', m = s) = \beta\,(1 - \delta_{s1}) b(s). \tag{10}$$

There is no double-counting problem here. A peculiar feature of the E-Z model is that the corresponding set of rate equations is semi-recursive, i.e. any k-th equation depends only on values of $n_{s'}$ for $s' \leq k$ and on a global constant depending on all n_s. This is a feature by which it is easy to show the existence and uniqueness of the solution and also to solve the system numerically. Both models assume that $a(s) = b(s)$. Fragmentation of the cluster triggered by any single member is described by $b(s) = s$

With the assumptions made so far, it turns out that each system is described by three constants: α, β and the total population size N. For the time-independent system we need just two constants, and since α and β are of dimension time^{-1} then only their ratio α/β should appear. Let us write explicitly the rate equations.

G-L system:

$$-\beta(s^2 - s)\, n_s - \alpha\, s\, n_s \sum_{s'=1}^{N} s' n_{s'} + \frac{\alpha}{2} \sum_{s'=1}^{s-1} s'\, n_{s'}\, (s - s') n_{s-s'} + \; 2\,\beta \sum_{s'=s+1}^{N} s'\, n_{s'} = 0\,. \tag{11}$$

E-Z system:

$$-\beta s\,(1 - \delta_{s1}) n_s - \alpha\, s\, n_s \sum_{s'=1}^{N} s' n_{s'} + \frac{\alpha}{2} \sum_{s'=1}^{s-1} s'\, n_{s'}\, (s - s') n_{s-s'} + \beta \delta_{s,1} \sum_{s'=s+1}^{N} s'^2\, n_{s'} = 0\,. \tag{12}$$

Eguíluz and Zimmermann[11] explicitly used the following constants:

$$\alpha = \frac{2(1 - \nu)}{N^2}\,, \qquad \beta = \frac{\nu}{N}\,. \tag{13}$$

We see that both sets of equations (11) and (12) simplify if we express them in terms of $k_s = s\, n_s$, i.e. the number of agents contained in clusters of size s. Note that for general $a(s)$, we need to substitute $k_s = a(s)\, n_s$.

B. Equilibrium in Gueron-Levin model: Continuous formulation

Gueron and Levin's solution [27] to the G-L model, was obtained for the system with continuous cluster density which we denote as $n(s)$. In terms of $k(s) = s\, n(s)$ the integral rate equation corresponding to Eq. (11) and with no limit on the maximum size of a cluster, is given by:

$$0 = -\beta\, s\, k(s) - \alpha\, k(s) \int_0^\infty ds'\, k(s') + \alpha \frac{1}{2} \int_0^s ds' k(s')\, k(s-s') + 2\beta \int_s^\infty ds'\, k(s') \,. \tag{14}$$

Looking at this equation we might guess that the solution is obtained by substituting an ansatz which satisfies $k(s+s') \propto k(s)k(s')$. The first form to try is $k(s) = A\, e^{-\mu\, s}$. With this ansatz we obtain

$$0 = -A\, \beta\, s\, e^{-\mu s} - A^2 \alpha/\mu e^{-\mu s} + A^2 \alpha/2\, s e^{-\mu s} + 2A\beta/\mu\, e^{-\mu s} \,. \tag{15}$$

There are two types of terms, either of the type $\sim\ e^{-\mu s}$ or $\sim\ s\, e^{-\mu s}$. Eliminating the overall exponential factor we have

$$0 = s \left(-A\beta + A^2 \frac{\alpha}{2} \right) + \frac{2}{\mu} \left(A\beta - A^2 \frac{\alpha}{2} \right) \,. \tag{16}$$

Both terms in parentheses have to vanish and this is indeed the case if we choose

$$A = 2\frac{\beta}{\alpha} \,. \tag{17}$$

The scale factor μ in the exponent is determined by normalization as $\mu = {}^{2\beta}/_{N\alpha}$. The solution to Eq. (14) is just an exponential function which was obtained by Gueron and Levin by means of a Laplace transform.

We notice here a remarkable curiosity: If we take the actual solution of Eq. (14), then for any s the following equalities hold exactly:

$$L_F(s) = G_C(s), \qquad L_C(s) = G_F(s) \,. \tag{18}$$

In other words, the following holds for the G-L model: *The average loss of clusters of size s due to the cluster fragmentation, is equal to the average gain obtained from the coalescence of clusters of sizes smaller than s. Also the average loss of clusters of size s due the coalescence with other clusters is equal to the average gain obtained from the fragmentation of clusters of sizes larger than s.* In addition to its mathematical interest, this identity (which is not satisfied for the E-Z model as discussed below) shows up a fundamental feature of the G-L model, which arises in turn from the microscopic rules which characterize it.

This symmetry also reveals if we look at behavior of the system with time flowing backwards (although we do not obtain a stochastic system by the time-reversal of recorded history of another non-equilibrium stochastic system, what becomes as an issue for discrete systems due to the presence of fluctuations, we may still discuss it considering the average quantities as describing the equilibrium state). With the reversed time perspective coalescence of clusters is observed as fragmentation and vice-versa, the average cluster size distribution is unaltered in equilibrium condition. We may draw then new coalescence/fragmentation rules which in case of G-L model are the same as original ones.

C. Cluster size distribution: The exponential cutoff

We now return to the discrete formulation. For the discrete version of the G-L system, it may be verified by direct computation that

$$n_s = 2\frac{\beta}{\alpha}\, s^{-1} \exp\left(-\mu\, s\right) \tag{19}$$

is also a solution of Eq. (11), once we make an approximation of extending the summation limits to infinity. Here, the normalization condition is $N = \sum_{s'=1}^{\infty} s'\, n(s')$, from which we calculate

$$\mu = \ln\left(\frac{2\beta}{\alpha N} + 1\right) . \tag{20}$$

Thus we have

$$n_s = 2\frac{\beta}{\alpha}\, s^{-1} \left(\frac{2\beta}{\alpha N} + 1\right)^{-s} . \tag{21}$$

It is advantageous to consider $\beta/\alpha \propto N$, thus the exponent is independent of N and n_s is just proportional to N. If we use here the same constants (Eq. (13)) as the original E-Z model, the solution is

$$\text{G-L}: \; n_s = N\frac{\nu}{1-\nu}\, s^{-1} \, (1-\nu)^s . \tag{22}$$

The solution to the E-Z model rate equations may be approximated as [29]

$$\text{E-Z}: \; n_s \sim N s^{-2.5} \left(\frac{4(1-\nu)}{(2-\nu)^2}\right)^s . \tag{23}$$

To compare the cluster size distribution for both models, we need to characterize both of them by the same parameters N and ν, i.e. they have the same coalescence function and their fragmentation functions agree for splitting of clusters of size $s = 2$. The cluster size distribution for both models is of the form $n_s \propto s^{-\kappa}\, e^{-\mu s}$. The scale of s at which the exponential cut-off becomes relevant may be defined if we look at the ratio

$$\frac{n_{s+1}}{n_s} = e^{-\mu}\, \frac{(s+1)^{-\kappa}}{s^{\kappa}} = e^{-\mu} \left(1 - \frac{\kappa}{s} + O\left(\frac{1}{s^2}\right)\right) . \tag{24}$$

The exponential cutoff becomes dominant at the scale when $a \approx (1 - \frac{\kappa}{s})$, hence we may define

$$s_{\text{cutoff}} \equiv \frac{\kappa}{1 - e^{-\mu}} \,. \tag{25}$$

For the models of interest in this paper, we have

$$\text{G-L}: \; s_{\text{cutoff}} = \frac{1}{\nu}\,, \quad \text{E-Z}: \; s_{\text{cutoff}} = \frac{5}{2}\left(\frac{2-\nu}{\nu}\right)^2 \,. \tag{26}$$

It is clear (see Fig. 3) that the range of cluster sizes for which one observes the power-law, is several orders of magnitude larger for the E-Z model than for the G-L model. We may also verify

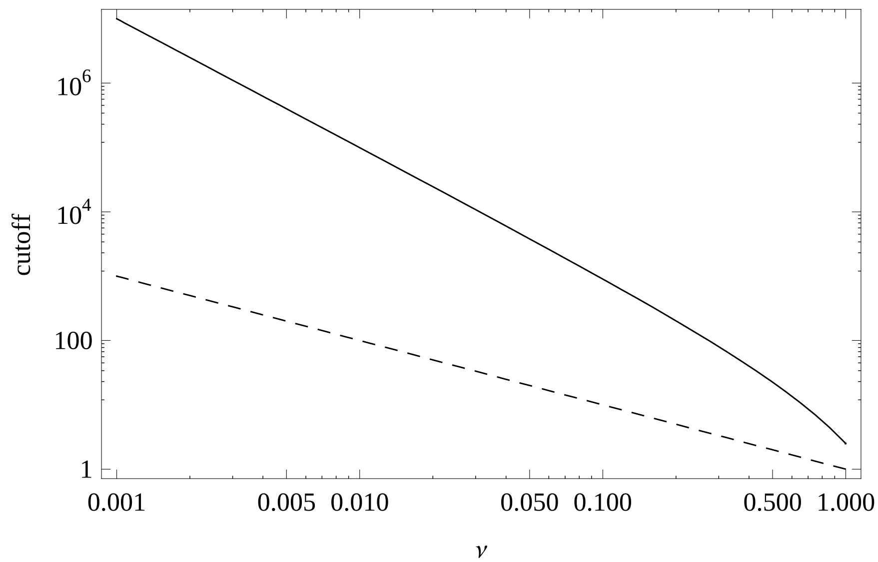

FIG. 3: Scale of exponential cutoff for the E-Z model (solid curve) and for the G-L model (dashed curve) described by the same parameter ν. The range of cluster sizes for which one observes the power-law, is several orders of magnitude larger for the E-Z model than for the G-L model.

that the special equilibrium result mentioned earlier for the continuous G-L model (see statement in italics), is also a property of the corresponding discrete model, once the upper limits in the sums are extended to infinity. It also holds that

$$\text{G-L model}: \; L_F(s) \cong s\frac{\nu}{1-\nu}L_C(s)\,, \quad \text{E-Z model}: \; L_F(s) = \frac{\nu}{2(1-\nu)}L_C(s)\,. \tag{27}$$

We can see therefore that for the G-L model, we can always find a value of s for which $L_F(s) \approx L_C(s)$ – in particular, it is the scale of the cluster size over which the exponential cutoff becomes apparent. By contrast, in the E-Z model for $\nu^{-1} \gg 1$ (i.e. for the large range of power-law

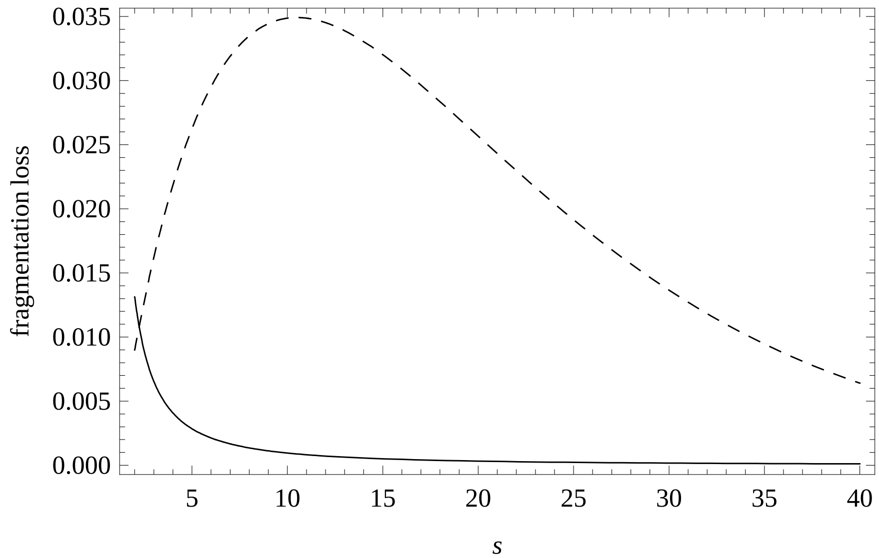

FIG. 4: $L_F(s)$, *loss due to fragmentation* for the E-Z model (solid curve) and for the G-L model (dashed curve) with parameters $\nu = 0.1$ and $N = 1000$. The overall scale is determined up to a multiplicative constant (i.e. the scale of time). The graphs show that $L_F(s)$ for the G-L model is usually much larger than $L_F(s)$ for the E-Z model.

behavior) we have $L_F(s) \ll L_C(s) \approx G_C(s)$. If we again compare both models, we find that $L_F(s)$ for the G-L model is usually much larger than $L_F(s)$ for the E-Z model. Figure 4 illustrates this finding for a particular set of parameters.

V. GENERALIZATION OF THE E-Z MODEL

We now open up the above discussion to a broader class of coalescence-fragmentation models. We note that overall, the number of coalescence-fragmentation-type process which have been employed to describe physical, biological and social systems, is enormous [8, 11, 15, 16, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39]. Our limited goal here is to select particular examples which illustrate how realistic corrections to the models used to mimic several social phenomena or financial markets behavior ("fat-tail problems") affect the cluster size distribution.

A. Spontaneous cluster formation

The first generalization we analyze is an effect of the presence of an extra term in the rate equations, this term is introduced to represent the following situation: A small number of clusters are allowed to form spontaneously from the population rather than from the merging of two smaller clusters — in practice this is most simply viewed as spontaneous formation of clusters from previously single agents/clusters of unit size (the exact mechanism is unimportant).

Let γ_s represent the rate describing the formation of clusters of size s by the non-hierarchical method. The value of γ_1 is implicitly defined by the requirement that the size N of the population remains constant, i.e., $\sum_{s=1}^{\infty} s\gamma_s = 0$.

The rate equation is given by

$$\begin{aligned}\frac{\partial n_s}{\partial t} &= -\alpha s n_s \\ &\quad +\beta \sum_{r=1}^{s-1} r n_r \left(s-r\right) n_{s-r} \\ &\quad -2\beta s n_s \sum_{r=1}^{\infty} r n_r + \gamma_s\end{aligned}$$

for $s \geqslant 2$.

In the steady state this may be written as

$$s n_s = A\left(\beta \sum_{r=1}^{s-1} r n_r \left(s-r\right) n_{s-r} + \gamma_s\right),$$

where A is defined by

$$A = \frac{1}{\alpha + 2\beta \sum_{r=1}^{\infty} r n_r}.$$

We have been also using the approximation $\sum_{r=1}^{\infty} r n_r f[r] \approx N$ extending the values of n_s and the sums to infinity.

The generating function $g[y]$ is now introduced:

$$g[y] \equiv \sum_{r=2}^{\infty} r n_r f[r] y^r. \tag{28}$$

By taking the square of this function and using Eq. (28) one obtains

$$\begin{aligned}0 &= \left(g[y]\right)^2 - \left(\frac{1}{A\beta} - 2n_1 y\right) g[y] \\ &\quad + n_1^2 y^2 + \frac{1}{\beta}\chi[y],\end{aligned} \tag{29}$$

where $\chi[y] \equiv \sum_{r=2}^{\infty} \gamma_r y^r$.

Using the fact that $g[1] = \sum_{r=1}^{\infty} r n_r - n_1$, this gives

$$n_1 = \frac{1 - A^2\left(\alpha^2 + 4\beta\chi[1]\right)}{4A\beta^{-1}}. \tag{30}$$

Solving Eq. (29) for general y and expanding the resulting radical using Taylor's theorem yields

$$g[y] \;=\; A\chi[y] + \frac{1}{2A\beta}\sum_{k=2}^{\infty}\left(\frac{(2k-3)!!}{(2k)!!} \times \left[4A\beta\left(n_1 y + A\chi[y]\right)\right]^k\right). \tag{31}$$

Here it will be assumed that the gamma term is small enough to be treated as a perturbation — specifically that $\frac{A\chi[y]}{n_1 y} \ll 1$, and thus a first-order binomial expansion of the exponential term in Eq. (31) may be performed. In this case:

$$g[y] \;\approx\; A\chi[y] + \frac{1}{2A\beta}\sum_{k=2}^{\infty}\left(\frac{(2k-3)!!}{(2k)!!} \times\left\{\left[4A\beta n_1 y\right]^k + k\frac{A}{n_1}\left[4A\beta n_1\right]^k \sum_{r=2}^{\infty}\gamma_r y^{r+k-1}\right\}\right).$$

Comparing terms in this with Eq. (28), after manipulation of the final summation, yields

$$n_2 = \frac{1}{2}A\gamma_2 + \frac{1}{2}A\beta(n_1)^2$$

for $s = 2$, and using Stirling's approximation one obtains, for large s,

$$n_s \;\approx\; \left(\frac{e^2}{4\sqrt{\pi}A\beta}\right)\left\{1 - A^2\left[\alpha^2 + 4\beta\mathrm{X}\right]\right\}^s s^{-5/2} + A\left[\gamma_s + \frac{e^2}{2\sqrt{\pi}n_1}\sum_{r=2}^{s-1}\left(\left\{1 - A^2\left[\alpha^2 + 4\beta\mathrm{X}\right]\right\}^r r^{-1/2}\gamma_{s-r+1}\right)\right]s^{-1}, \tag{32}$$

where $\mathrm{X} \equiv \chi[1]$.

Since A is constant for a given population the general form of the above equation is

$$n_s \approx \Xi\kappa^s s^{-5/2} + Z[s]s^{-1}, \tag{33}$$

where $Z[s]$ describes the details of the perturbation.

B. Step perturbation

Hereafter we study the example of a highly idealized perturbation which is intended to imitate a more general form in which the perturbation is reasonable at small s and dies off as s increases.

The simplest form is a step function:

$$\gamma_s = \begin{cases} \frac{\Phi}{q-1}, & \text{for } 2 \leqslant s \leqslant q; \\ 0, & \text{for } s > q; \end{cases}$$

where q is an arbitrarily chosen cluster size and $\Phi > 0$.

Using the original E-Z parametrization 13 and Eq. 32 we obtain the cluster size distribution as

$$n_1 \approx N\frac{1-\Phi}{2-\nu},$$

$$n_2 \approx \frac{1}{2(2-\nu)}\left[\frac{\Phi}{q-1} + \frac{1-\nu}{(2-\nu)^2}(1-\Phi)^2\right],$$

$$\begin{aligned} n_s \;\approx\; & N\frac{(2-\nu)e^2}{4\sqrt{\pi}(1-\nu)}\left[\frac{4(1-\nu)}{(2-\nu)^2}(1-\Phi)\right]^s s^{-5/2} \\ & + N\frac{1}{2-\nu}\frac{\Phi}{q-1}s^{-1} \\ & + \frac{e^2}{2\sqrt{\pi}}\frac{\Phi}{(1-\Phi)}\frac{1}{q-1} \\ & \times\left\{\sum_{r=2}^{s-1}\left[\frac{4(1-\nu)}{(2-\nu)^2}(1-\Phi)\right]^r r^{-1/2}\right\} s^{-1}, \end{aligned}$$

for $3 \leqslant s \leqslant q$, and

$$\begin{aligned} n_s \;\approx\; & N\frac{(2-\nu)e^2}{4\sqrt{\pi}(1-\nu)}\left[\frac{4(1-\nu)}{(2-\nu)^2}(1-\Phi)\right]^s s^{-5/2} \\ & + \frac{e^2}{2\sqrt{\pi}}\frac{\Phi}{(1-\Phi)}\frac{1}{q-1} \\ & \times\left\{\sum_{r=2}^{q}\left[\frac{4(1-\nu)}{(2-\nu)^2}(1-\Phi)\right]^r r^{-1/2}\right\} s^{-1} \end{aligned}$$

for $s \geqslant q+1$.

The examples of predicted distribution of n_s are plotted in Fig. 5. It is interesting to note here, that the greatest deviation is found at high s, whereas the perturbation, in its nature, only directly affects clustering at low s. This seems likely to be due to the effective nucleation of larger clusters by the perturbation — by its creation of small clusters by non-hierarchical means, the perturbation greatly accelerates the formation of large clusters even though the small clusters still fragment at a sufficient rate to hide the effect at low s on the graph. In Fig. 6 the predicted distribution of n_s is

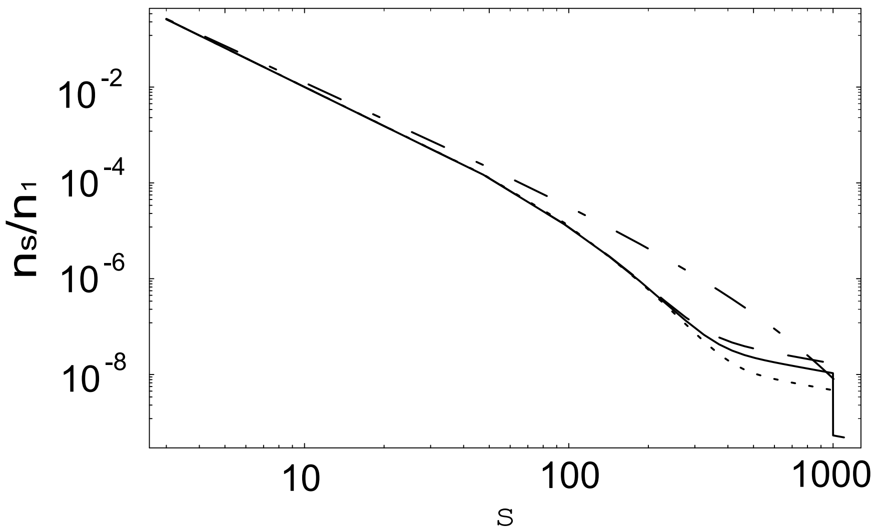

FIG. 5: Predicted distribution of cluster sizes in population of Section V B, using $N = 1000$, $\nu = 0.1$ and $\Phi = 0.01$. The dot-dashed line shows the unperturbed population; the dashed line shows $q = 10$; the dotted line shows $q = 100$; the solid line shows $q = 1000$.

plotted for different signs of the perturbation ($\pm\Phi$) along the unperturbed prediction. For the case with the negative sign in order that n_s remain finite as $s \to \infty$ it is necessary that

$$\Phi < \frac{\nu^2}{4\,(1-\nu)}. \tag{34}$$

It can be seen that the deviations of the perturbed populations are as expected. The difference can perhaps best be seen in direct numerical comparison (the primed quantities refer to the $-\Phi$ case)– the model predicts the ratios below, for $N = 10000$, $\Phi = 0.001$, $\nu = 0.1$ and $q = 500$:

$$\begin{aligned} \frac{n_1}{n_1'} &= 0.998, \\ \frac{n_{500}}{n_{500}'} &= 0.39, \\ \frac{n_{1000}}{n_{1000}'} &= 0.14. \end{aligned}$$

This is a marked difference for such a small perturbation. (The interpretation of the perturbation is that statistically a cluster of size 500 or less spontaneously forms/fragments for $+/-\Phi$ cases respectively once in every 1000 timesteps, where by a timestep we mean any occurring fragmentation or coalescence in the system.)

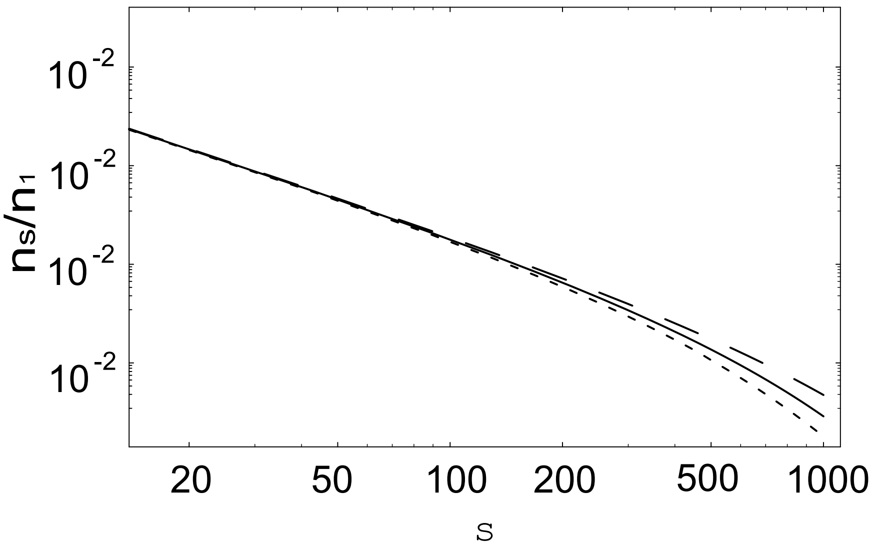

FIG. 6: Predicted cluster size distribution for the $+\Phi$ case (dotted) and the $-\Phi$ case B(dashed) compared with the unperturbed model (solid). Parameter values: $N = 10000$; $\Phi = 0.001$; $\nu = 0.1$; $q = 500$.

C. Variable population size

So far in this paper we have dealt purely with populations of some fixed size N. Evidently such a closed system is only applicable to a small subset of real-world problems. Hence in this section of the paper we develop an analytic treatment of a model which is analogous to the E-Z model, but treats the case of a population whose size varies with time according to a simple law.

Our model proceeds as follows, for a population containing $N[t]$ agents divided into $M[t]$ clusters in the style of the E-Z model. First we consider a case in which the change in the number of agents is positive.

1. In the time unit, with probability $p[t]$, $L[t]$ new agents are added to a single cluster of size s, the cluster being selected with probability proportional to s.

2. Alternatively, with probability $q[t] = 1 - p[t]$, a randomly selected cluster fragments (selection of this cluster is independent of cluster size).

If the change in the number of agents is negative, then the model runs as follows:

1. In the time unit, with probability $p[t]$, $L[t]$ agents are removed from a single cluster of size s, the cluster being selected with probability proportional to s. If the selected cluster has $s < |L[t]|$ then nothing occurs.

2. Alternatively, with probability $q[t] \equiv 1 - p[t]$, a randomly selected cluster fragments (selection of this cluster is independent of cluster size).

The rationale behind adding or subtracting from a *single* cluster is the conjecture that in the situations of interest, only a single cluster will be involved in an external event affecting the population.

1. *Increasing population size:* $L[t] > 0$

The model proposed above leads to the rate equations

$$\begin{aligned}
\frac{\partial n_s}{\partial t} &= \frac{p[t]}{N[t]}\left((s-L[t])n_{s-L[t]} - sn_s\right) - \frac{q[t]}{M[t]}n_s \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{for } s > L[t], \\
\frac{\partial n_s}{\partial t} &= \qquad -\frac{p[t]}{N[t]}sn_s - \frac{q[t]}{M[t]}n_s \qquad \text{for } 2 \leqslant s \leqslant L[t], \\
\frac{\partial n_1}{\partial t} &= \quad -\frac{p[t]}{N[t]}n_1 + \frac{q[t]}{M[t]}\textstyle\sum_{r=2}^{\infty} rn_r \qquad \text{for } s = 1,
\end{aligned}$$

with resulting totals

$$\frac{\mathrm{d}N}{\mathrm{d}t} = p[t]L[t], \tag{35}$$

$$\frac{\mathrm{d}M}{\mathrm{d}t} = q[t]\left(\frac{N[t]}{M[t]} - 1\right). \tag{36}$$

The solution of the above equations clearly depends on the forms of $L[t]$ and $p[t]$. As a simple example, we take both to be constant: $L[t] \equiv L$ and $p[t] \equiv p$ for all t. In this case, it can be seen that for times $t \gg \frac{N[t=0]}{pL}$, Eq. (35) yields the linear solution

$$N[t] = pLt.$$

If we assume a similar asymptotically linear form for $M[t]$ at large t, $M[t] = \sigma t$, we can go on to deduce from Eq. (36) that

$$\sigma = \frac{q}{2}\left(\sqrt{4\frac{p}{q}L + 1} - 1\right).$$

We now assume a linear form for all n_s: $n_s[t] = c_s t$. In this case, one obtains the solution

$$c_1 = \frac{q}{\sigma}\frac{L}{L+1}\sum_{k=1}^{\infty}(1+kL)c_{1+kL} = \frac{q}{\sigma}\frac{L}{L+1}\left(pL - c_1\right).$$

Therefore,

$$c_1 = \frac{pq}{\sigma}\frac{L^2}{L(1+q/\sigma)+1},$$
$$c_{1+kL} = \frac{pq}{\sigma}\frac{L^2}{L(1+q/\sigma)+1}\frac{\rho!^{(L)}(1+(k-1)L)!^{(L)}}{(\rho+kL)!^{(L)}}, \quad (37)$$

for $k = 1, 2, 3,\dots$, where

$$\rho \equiv \left(\frac{q}{\sigma}+1\right)L+1,$$

and we have used the multifactorial function, defined by

$$m!^{(n)} = \begin{cases} 1, & \text{if } 0 \leqslant m < n; \\ m(m-n)!^{(n)}, & \text{if } m \geqslant n. \end{cases}$$

Clearly $c_s = 0$ for $s \neq 1 + kL$.

Now via a generalization of Stirling's approximation, it can be shown that

$$\ln(n!^{(b)}) \sim \frac{1}{b}(n\ln n - n).$$

With this approximation applied to Eq. (37), we finally obtain our solution:

$$c_{1+kL} \approx \frac{pq}{\sigma}\frac{L^2}{L(1+q/\sigma)+1}\frac{e^{(L-1)/L}\rho^{\rho/L}(1+(k-1)L)^{k-1+1/L}}{(\rho+kL)^{k+\rho/L}} \quad (38)$$

for integer $k \geqslant 1$. If we take a snapshot of this system at any time the observed cluster size distribution will be given by Eq. (38) modulo multiplicative constant growing linearly with time. The leading $k-$behavior of Eq. (37) is

$$\frac{(kL+(1-l))^{1-L}}{(kL+\rho)^{\rho})}(kL)^{-(\rho+L-1)} \ . \quad (39)$$

2. *Decreasing population size:* $L[t] < 0$

In the following analysis, for simplicity's sake we do not allow complete annihilation of clusters (i.e., the removal of all of a cluster's members from the population).

The rate equations for $L[t] < 0$ are as follows:

$$\frac{\partial n_s}{\partial t} = \frac{p[t]}{N[t]}\left((s+|L[t]|)n_{s+|L[t]|} - sn_s\right) - \frac{q[t]}{M[t]}n_s \quad \text{for } s > |L[t]|,$$
$$\frac{\partial n_s}{\partial t} = \frac{p[t]}{N[t]}(s+|L[t]|)n_{s+|L[t]|} - \frac{q[t]}{M[t]}n_s \quad \text{for } 2 \leqslant s \leqslant |L[t]|,$$

and

$$\frac{\partial n_1}{\partial t} = \frac{p[t]}{N[t]}(1 + |L[t]|)n_{1+|L[t]|} + \frac{q[t]}{M[t]} \sum_{r=2}^{\infty} r n_r$$

for $s = 1$, with resulting totals

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\frac{p[t]|L[t]|}{N[t]} \sum_{r=1+|L|}^{\infty} r n_r, \tag{40}$$

$$\frac{\mathrm{d}M}{\mathrm{d}t} = q[t]\left(\frac{N[t]}{M[t]} - 1\right). \tag{41}$$

As above, to obtain a solution we assume that p and L are both constant, and proceed to try a linear solution of the form

$$N[t] = N_0 - \gamma t,$$
$$M[t] = M_0 + \sigma t,$$
$$n_s[t] = C_s - c_s t.$$

This approximation can only hold as long as the changes in each n_s are small compared in relation to the size of the respective C_s. In this case — i.e., for t not too large — we obtain

$$\sigma \approx q\left(\frac{N_0}{M_0} - 1\right),$$
$$\gamma \approx \frac{p|L|}{N_0} \sum_{r=1+|L|}^{\infty} r C_r,$$

and for the c_s:

$$c_1 \approx -\frac{p}{N_0}(1 + |L|)C_{1+|L|} - \frac{q}{M_0}(N_0 - C_1) \quad \text{for } s = 1,$$

$$c_s \approx \frac{q}{M_0}C_s - \frac{p}{N_0}(s + |L|)\,C_{s+|L|} \quad \text{for } 2 \leqslant s \leqslant |L|,$$

$$c_s \approx \left(\frac{q}{M_0} + \frac{p}{N_0}s\right)C_s - \frac{p}{N_0}(s + |L|)\,C_{s+|L|} \quad \text{for } s > |L|.$$

Therefore, with a suitable choice of initial conditions and a large population, one can infer the small-t behavior of the system.

3. Decreasing population: proof of concept

As a simple example case, we take $L < 0$ and a starting population of the form

$$n_s[t=0] = \begin{cases} C_1 - \phi s, & \text{if } s < \frac{C_1}{\phi}; \\ 0, & \text{if } s \geqslant \frac{C_1}{\phi}. \end{cases}$$

In this case our equations from Section V C 2 yield

$$\begin{aligned} N_0 &= \frac{1}{6} C_1 \left[\left(\frac{C_1}{\phi} \right)^2 - 1 \right], \\ M_0 &= \frac{1}{2} C_1 \left(\frac{C_1}{\phi} - 1 \right), \\ \gamma &\approx p\,|L| \left\{ 1 - \frac{|L|\,(1+|L|)\,(3C_1 - \phi - 2|L|\phi)}{C_1 \left[\left(\frac{C_1}{\phi} \right)^2 - 1 \right]} \right\}, \\ \sigma &\approx \frac{(1-p)}{3} \left(\frac{C_1}{\phi} - 2 \right). \end{aligned}$$

This leads to an expression for n_1 of the form

$$\begin{aligned} n_1[t] \approx C_1 + \bigg[& \frac{p}{N_0} (1+|L|) (C_1 - \phi (1+|L|)) \\ & + \frac{q}{M_0} (N_0 - C_1) \bigg] t, \end{aligned}$$

with corresponding n_s of the form

$$\begin{aligned} n_s[t] \approx C_1 - \phi s - \bigg[& \left(\frac{q}{M_0} C_1 - \frac{|L| p}{N_0} C_{|L|} \right) \\ & - \left(\frac{p}{N_0} C_{2|L|} + \frac{q}{M_0} \phi \right) s + \frac{p}{N_0} \phi s^2 \bigg] t \end{aligned}$$

for $2 \leqslant s \leqslant |L|$, and

$$\begin{aligned} n_s[t] \approx C_1 - \phi s - \bigg[& \left(\frac{q}{M_0} C_1 - \frac{|L| p}{N_0} C_{|L|} \right) \\ & + \left(\frac{2\,|L|\,p}{N_0} - \frac{q}{M_0} \right) \phi s \bigg] t \end{aligned}$$

for $s > |L|$. Figure 7 shows a plot of this model using reasonable parameter values.

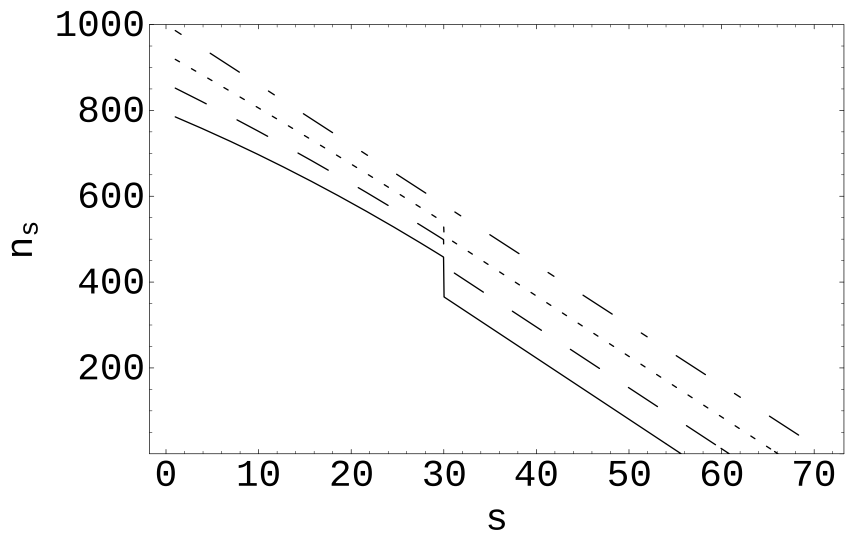

FIG. 7: The predictions of the model of Section V C 3, using parameter values $C_1 = 1000$, $\phi = 14$, $p = 0.3$ and $L = -30$. This yields population size parameters of $N_0 = 850173$ and $M_0 = 35214$. Line styles reflect different values of the parameter t: dot-dashed$(t = 0)$, dotted$(t = 5000)$, dashed$(t = 10000)$ and solid$(t = 15000)$. Beyond $t = 15000$ it can be seen that the approximations made in the derivation of Section V C 2 become inaccurate.

D. Further real-world generalizations of the Eguiluz-Zimmermann model

We now open up the above discussion to a broader class of coalescence-fragmentation models. The framework described in this chapter provides a common basis for discussing these more general models. Some of these more general models have been published before and some have not. We note that overall, the number of coalescence-fragmentation-type process which have been employed to describe physical, biological and social systems, is enormous [8, 11, 15, 16, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39]. Our limited goal here is to select particular examples which illustrate how realistic modifications of the coalescence-fragmentation function affect the cluster size distribution. Note that in this section, the models follow similar rules to the E-Z model in that they exhibit binary coalescence and total fragmentation, unless otherwise specified. As we shall see, small changes in the coalescence-fragmentation rules can sometimes yield dramatic changes in the cluster size distribution, and vice versa. In other words, the 'devil is in the detail' in terms of the emergent phenomena that can be expected from a given set of microscopic rules. Such specificity is important to establish, since any particular stylized facts which are observed empirically in the macroscopic behavior might then be traceable to specific sets of microscopic rules – or at the very least, may help to reduce the number of possible candidate rules. We start by looking at variants in which the total population size N is treated as one of the parameters defining the model. Specifically, we consider a constant population such that the constraint $N = \sum_{s=1}^{N} s n_s(t)$ holds.

Fragmentation into clusters of fixed size: In this scenario, any cluster may only fragment into a set of smaller clusters, each of fixed size s_0. For the discrete system there is naturally a divisibility problem regarding fragmentation of clusters of sizes which are not a multiple of s_0. Since we are interested in steady-state behavior, we may assume that such clusters do not fragment. Whatever the initial configuration is after a sufficiently long time, the system in equilibrium will consist almost entirely of clusters that are a multiple of s_0 in size. It turns out that the cluster size distribution has the same form as the E-Z model in Ref. (23), if we re-express it in terms of s_0 as the basic unit, i.e. if we substitute $s \rightarrow s/s_0$.

Scale-dependent behavior of clusters: In Ref. [31], the fragmentation rate $f(s)$ and coalescence function $\mathcal{C}(s, s')$ take the following form which depends on cluster size:

$$f(s) = \begin{cases} p(s) & s \leq s_0 \\ q(s) & s > s_0 \end{cases} \tag{42}$$

$$\mathcal{C}(s, s') = \begin{cases} p(s)p(s'), & s \leq s_0 \;\; s' \leq s_0 \\ p(s)q(s'), & s \leq s_0 \;\; s' > s_0 \\ q(s)p(s'), & s > s_0 \;\; s' \leq s_0 \\ q(s)q(s'), & s > s_0 \;\; s' > s_0 \end{cases} \tag{43}$$

where s_0 represents some characteristic cluster size, $p(s) = (\frac{s}{s_0})^{\delta}$ and $q(s) = (\frac{s_0}{s})^{\delta}$. Two different scaling behaviours are observed:

$$n_s \sim \begin{cases} N s^{-(\frac{5}{2}+\delta)}, & s \leq s_0 \\ N s^{-(\frac{5}{2}-\delta)}, & s > s_0 \end{cases}$$

Another related case [32] is where $p(s) = q(s) \sim s^{-\delta}$ and $s_0 = 1$, in which case $n_s \sim s^{-(5/2-\delta)}$.

Constant fragmentation probability: The model proposed in Ref. [30] employs a fragmentation scenario in which a cluster that is selected with probability proportional to s will fragment with probability $\nu\gamma/s$, while with probability $\nu(1-\gamma/s)$ the selected cluster does nothing. Introducing timesteps where nothing happens simply changes the value of the time scale but does not influence the time-independent behavior. The rate that the cluster fragments into a bunch of single members is therefore constant. If it is small enough according to Ref. [30] this system exhibits scale-free behavior, but with a power-law exponent which depends on the parameters specifying the model: $n_s \sim As^{-(\kappa+1)}$, where $\kappa = 2/(1 - \sqrt{1 - 4/\alpha})$ and $\alpha = 2\nu\gamma/(1-\nu)$. The power-law is observed for $\alpha > 4$ which corresponds to $\kappa \geq 2$.

Heterogeneity of members: In many real-world systems, especially in biological or social models, we encounter a heterogeneous population. In Ref. [36], a character is introduced as an m-dimensional normalized vector, formed from m-bit binary strings. Their scalar product then becomes the argument of a function which controls the coalescence and fragmentation processes. The general case requires numerical simulation. Interestingly, however, this model produces a power-law over part of its range with a slope identical to the E-Z model. It is the form of the exponential cut-off, but not the exponent itself, which depends on the heterogeneity of the population. We recently explored another type of heterogeneous E-Z-like model, showing that it can bridge the gap between the power-law slope of magnitude 2.5 for clusters in the E-Z model (and hence 1.5 for price returns) and the empirical value of financial market price returns which is typically closer to 4 [20]. A simple version of the vector model is provided via a fascinating recent variation proposed by Hui[37] in which the heterogeneity is represented by a character parameter $\rho_k \in [0, 1]$ which is assigned to each object in the entire population, where objects are numbered by $k = 1 \ldots N$. The probability that an agent i and another agent j to form a link (and therefore for the inequivalent clusters to which these members belong to merge) depends on the value $|\rho_i - \rho_j|$. In principle it may be a general symmetric function $p(\rho_i - \rho_j)$. The fragmentation of a cluster may also depend on the characters of the members that form the particular cluster. One way of introducing this is by a mechanism in which fragmentation of the whole cluster is triggered by breaking any single link that belongs to it [37]. Since a weaker link is easier to break, it is assumed that the probability that the link breaks is proportional to $p(\rho_i - \rho_j)$ which may be interpreted as a measure of the strength of the link formed between members i and j. If $p(\rho_i - \rho_j)$ is a function which is sharply peaked at 0, we will have a situation where the newly formed clusters consist only of members of very similar character, and the whole system may be considered as a mixture of several homogeneous population subsystems which do not interact which each other. Each of these subsystems is described by the cluster size distribution of the form in Eq. (23) with constants determined by the distribution of characters across the population. The cluster size distribution for the whole system (regardless of the character) is then a sum of the distributions for the subsystems, therefore we still observe a scale-free behavior with variation in the form of the cut-off (i.e. diversity in the heterogeneity of the population induces diversity in the constants describing the subsystems and lengthens the tail of the size distribution tail). In the opposite limiting case, the function $p(\rho_i - \rho_j)$ does not vary sharply over its argument, e.g. $p(\rho_i - \rho_j) \propto 1 - |\rho_i - \rho_j|$. In such cases we encounter homogeneous mixing – the distribution of characters across different clusters is uniform and the system might be described as a homogeneous one by Eqs. (12) and (23). The presence of the heterogeneity changes only the value of α/β in Eq. (12).

Reservoir model: This the case of a non-conserved population, the total population size is no longer a parameter defining the model, but becomes a dynamical variable whose average value in equilibrium is determined by the model itself: $N \equiv \sum_{s=1}^{\infty} s\langle n_s(t)\rangle$ We introduce a constant supply of individuals from a system reservoir, with γ denoting the rate in which single individuals are added. The products of the fragmenting cluster are then moved back to the reservoir. An equivalent interpretation is that a cluster stays in the system but ceases to interact (i.e. it does not merge with other clusters). The remainder of the dynamics resembles the terms in the E-Z model, with βs being the probability rate of removing a cluster of size s and $\alpha s s'$ being the coalescence rate. This particular reservoir model is therefore described by three parameters α, β, γ, with only two parameters required for the steady-state cluster size distribution. The master equations are

$$- \beta s\, n_s - \alpha\, s\, n_s \sum_{s'=1}^{\infty} s' n_{s'} + \frac{1}{2}\, \alpha \sum_{s'=1}^{s-1} s'\, n_{s'}\, (s - s') n_{s-s'} + \gamma \delta_{s1} = 0\,. \tag{44}$$

By summation of Eq. (44), the average number of participants is obtained as

$$\langle N \rangle = \frac{\sqrt{\beta^2 + 2\alpha\gamma} - \beta}{\alpha}\;. \tag{45}$$

The cluster size distribution has the same form as for the E-Z model Eq. (23), if expressed in terms of $\langle N \rangle$ and α/β. In this case there is no approximation made in extending the summation limit to infinity, and the solution in Eq. (23) is exact from the mean-field theory point of view. There is no limit on the maximum size of a cluster, which in principle may exceed $\langle N \rangle$ when the effect of fluctuations is non-negligible.

VI. CONCLUSIONS AND IMPLICATIONS

We have examined a wide variety of coalescence-fragmentation systems, and have pointed out how subtle changes in their underlying rules can affect the resulting distribution of cluster sizes. In the process, we have managed to connect the rules of coalescence and fragmentation with terms in the corresponding rate equations, and have identified the specific ways in which they affect the resulting distribution of cluster sizes. The connections are not always direct, but we have offered various insights which help establish a more direct link. In each case studied, the system senses the fragmentation function in two ways: the appearance of new clusters coming from the fragments of the fragmented cluster (represented by $G_F(s)$), and the disappearance of clusters that fragment (represented by $L_F(s)$).

As a result of our analysis, we can better understand what factors dictate when a power-law is likely to emerge, and what tends to control its exponent. We conclude that: (1) it is the substantial contribution of $L_F(s)$ in the equilibrium condition (Eq. 1) which may prevent the size distribution from showing a power-law behavior. (2) The presence or absence of $G_F(s)$ (i.e. the appearance of fragmentation products of new clusters) influences strongly the value of the power-law exponent itself, in cases where the power-law emerges. In the case when the parameter controlling the fragmentation is small but finite, it is hard to identify a common limiting case for the various systems studied – however, the form of the fragmentation function does influence the cluster size distribution regardless of the value of this parameter. Note that if the fragmentation rate tends to zero, the system cannot be clearly described by mean-field theory, since it performs quasi-oscillatory behavior associated with building up one supercluster containing the whole population. This supercluster then eventually breaks up. Whatever the mode of fragmentation, the exponent of the power-law may be controlled by altering the power of the cluster size s which is involved in the fragmentation and coalescence function. Specifying it realistically requires some understanding at the microscopic level. The most common mechanism of coalescence is created by building random links between the population members, yielding a coalescence function of the form $\sim ss'$.

If we adopt a point of view in which the system is considered as an evolving network, the clusters represent disconnected components. Depending on the particular rules, the fragmentation process now corresponds to breaking links. If the disconnected component in a network breaks predominantly into single members, it might be still interpreted in terms of the fragmentation being triggered by a single member, provided we allow some kind of link-breaking virus to spread rapidly throughout the entire disconnected component. Somewhat counter-intuitively, we have also seen that the behavior of the heterogeneous system does not substantially differ from the behaviour of the homogeneous one. This results from two effects: the homogeneous mixing effect, and the coexistence of several non-interacting populations whose distinct ‘characters’ lie hidden in the cluster size distribution.

Although we have mentioned various possible applications, we finish by noting a new one. Many of the neurodegenerative disorders associated with aging, for example Alzheimer's disease, are thought to be associated with the large-scale self-assembly of nanoscale protein aggregates in the brain [39]. Protein-aggregation has of course attracted much attention over the years in both the chemistry and physics literature – however, the problem of protein aggregates in neurodegenerative diseases is known to be much harder than traditional polymer problems, because of the complexity of the individual proteins themselves [39]. Given the wide range of possible heterogeneities *in vivo* within a cell, there is typically insufficient knowledge to specify either (i) a specific diffusion model and its geometry and boundary conditions, as a result of geometrical restrictions and crowding effects[18], or (ii) a specific reaction model for the binding rates, given the wide variety of conformational states in which molecules may meet. It therefore makes sense to assign some probabilities to the aggregation process – and in particular, coalescence and fragmentation probabilities to describe the joining of an n-mer with an n'-mer to give an n''-mer, where $\{n, n', n''\} \equiv 1, 2, 3, ...$, and its possible breakup. The precise details of the coalescence and fragmentation rules now takes on a critical importance, since subtle changes in these rules can alter the resulting size distribution of the n-mer population. The practical question of how fatal a given realization of the disease will be in a particular patient, becomes intertwined with the question of whether the distribution of cluster sizes is a regular one in terms of its fluctuations – e.g. a Gaussian or Poisson distribution which both have a finite variance – or it is a power-law which may then have an infinite variance. Although in practice a cut-off always exists, a power-law with an exponent $\alpha < 2$ has (in principle) an infinite mean and infinite standard deviation; a power-law with $2 < \alpha < 3$ has (in principle) a finite mean but an infinite standard deviation; and a power-law with $\alpha > 3$ has a finite mean and finite standard deviation. The implication is that a coalescence-fragmentation process producing a power-law with $\alpha < 3$ as in E-Z-type models where $\alpha \sim 2.5$, has a significant probability of forming very large n-mers because of its (in principle) infinite standard deviation. Suppose for the moment that an n-mer of size $n \geq n_0$ can produce a neurodegenerative disorder, then the fraction of such dangerous n-mers in a soup of self-assembling polymer aggregates, will be non-negligible if $\alpha < 3$. In the highly crowded, heterogeneous n-mer population expected in the human body, the resulting value of any approximate power-law slope α could therefore be a crucial parameter to estimate. The possibility of engineering this α value such that large aggregates are unlikely, through subtle changes in the coalescence and fragmentation processes, then takes on a very real possibility and adds some direct medical relevance to this work.

---

[1] M. Smoluchowski: Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Klloidteilchen. *Physik. Zeitschr.*, **17**, 557-599 (1916).

[2] M. Smoluchowski: Versuch einer mathematicschen Theoie der Koagulationskinetik kolloider Lösungen. *Zeitschrift f. physik. Chemie,* **92**, 129-168 (1917).

[3] F. Leyvraz: Large-time behavior of the Smoluchowski equations of coagulation, Phys. Rev. A **29**, 854 - 858 (1984)

[4] F. Leyvraz: Scaling Theory and Exactly Solved Models In the Kinetics of Irreversible Aggregration. *Phys. Rep.* **383**, 95 (2003).

[5] Ph. Laurençot, On a class of continuous coagulation-fragmentation models, *J. Differential Equations* **167**, 145-174 (2000)

[6] Ph. Laurençot, S. Mischler, Global existence for the discrete diffusive coagulation-fragmentation equation in L1, *Rev. Mat. Iberoamericana* **18**, 731-745 (2002)

[7] P. Laurençot and S. Mischler: On coalescence equations and related models. In *Modeling and computational methods for kinetic equations*, Model. Simul. Sci. Eng. Technol., 321-356. Birkhäuser Boston, Boston, MA, (2004).

[8] J. Ferkinghoff-Borg, M. H. Jensen, J. Mathiesen, and P. Olesen: Scale-free cluster distributions from conserving merging-fragmentation processes Europhys. Lett., **73**(3), 422 428 (2006).

[9] R. L. Drake: A general mathematical survey of the coagulation equation. In G. M. Hidy and J. R. Brock, editors, *Topics in Current Aerosol Research (Part 2)* , volume 3 of *International Reviews in Aerosol Physics and Chemistry*, 201-376. Pergamon, (1972).

[10] D. J. Aldous: Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernouilli*, 5(1):3-48, (1999)

[11] V. M. Eguíluz and M. G. Zimmermann: Transmission of Information and Herd Behaviour: An Application to Financial Markets. Phys. Rev. Lett. **85**, 5659–5662 (2000).

[12] J. M. Ball, J. Carr, The discrete coagulation-fragmentation equations:existence, uniqueness and density conservation, J. Stat. Phys. **61**, 203-234 (1990)

[13] J. Carr, Asymptotic behavior of solutions to the coagulation-fragmentation equations. I. The strong fragmentation case, Proc. Roy. Soc. Edinburgh Sect. A, **121**, 231-244 (1992)

[14] J. Carr, F. P. da Costa, Asymptotic behavior of solutions to the coagulation-fragmentation equations. II. Weak fragmentation, J. Stat. Phys. **77**, 89-123 (1994)

[15] N. F. Johnson, Policing (2008) doi: 10.1093/police/pan018; N. F. Johnson, in *Managing Complexity: Insights, Concepts, Applications* ed. D. Helbing (Springer, Berlin, 2008) p. 303. See also Neil F. Johnson, Mike Spagat, Jorge A. Restrepo, Oscar Becerra, Juan Camilo Bohorquez, Nicolas Suarez, Elvira Maria Restrepo, Roberto Zarama, *Universal patterns underlying ongoing wars and terrorism*, LANL e-print arXiv:physics/0605035.

[16] N. F. Johnson, P. Jefferies and P.M. Hui: Financial Market Complexity (Oxford University Press, 2003).

[17] P. M. Hui, private communication.

[18] S. Schnell and R. Hancock, *The intranuclear environment* in Methods in Molecular Biology - The Nucleus, to appear (2008).

[19] *Large Scale Structures and Dynamics of Complex Networks*, ed. G. Caldarelli and A. Vespignani (World Scientific, Singapore, 2007).

[20] A. Kirou, B. Ruszczycki, M. Walser and N.F. Johnson: Computational Modeling of Collective Human Behavior: The Example of Financial Markets. in *Proceedings of the International Conference on Computational Science 2008* (Springer, Berlin, 2008) p. 33 in part 1.

[21] See for example, the wide range of publications and conferences around this common theme of computational modeling of socio-economic systems, available at website *www.unifr.ch/econophysics*.

[22] J.-P. Bouchaud, M. Potters: Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management. Cambridge University Press, Cambridge, Second Edition (2004).

[23] R. N. Mantegna, H. E. Stanley: An Introduction to Econophysics: Correlations and Complexity in Finance. Cambridge University Press, Cambridge (1999).

[24] N. F. Johnson: Two's company, three is complexity. Oneworld, New York, (2007).

[25] R. Cont, J.-P. Bouchaud: Herd Behavior and Aggregate Fluctuations in Financial Markets. Macroeconomic Dynamics **4**, 170–196 (2000).

[26] D. Challet and Y.C. Zhang:Emergence of Cooperation and Organization in an Evolutionary Game. Physica A **246**, 407–418 (1997).

[27] S. Gueron and S.A. Levin: The Dynamics of Group Formation. Mathematical Biosciences 128, 243–246 (1995).

[28] Y.B. Xie, B.H. Wang, H.J. Quan, W.S. Yang and P.M. Hui, Physical Review E, Volume 65, 046130

(2002).

[29] R. D’hulst and G.J. Rodgers, International Journal of Theoretical and Applied Finance 3 609 (2000) cond-mat/9908481.

[30] Y.B. Xie,B. Hu and T.Zhou, Physical Review E 71, 046135 (2005).

[31] D.F. Zhang, G.J. Rodgers, P.M. Hui, Physica A 310, 480–486 (2002).

[32] D.F. Zhang, P.M. Hui, K.F. Yip and N.F. Johnson, Eur. Phys. J. B 27, 213–218 (2002).

[33] G.J. Rodgers and D.Zheng, Physica A 308, 375–380 (2002).

[34] G.J. Rodgers and Y.J. Yap, Eur. Phys. J. B 28, 129–132 (2002).

[35] S. Rawal,G.J. Rodgers, Physica A 344, 50–55 (2004).

[36] A. Wyld, G.J. Rodger, Physica A 374, 491–500 (2007).

[37] P. M. Hui, private communication.

[38] E. Bonabeau, L. Dagorn and P. Freon, 1999 Proc. Natl Acad. Sci. USA 96, 4472–4477. (doi:10.1073/pnas.96.8.4472).

[39] A.J. Modler, K. Gast, G. Lutsch, and G. Damaschun, J. Mol. Biol. 325, (2003).

[40] V.M. Kolybasova and Yu.N. Sokolskikha, Physics Letters B 276, 409–412 (1992)

[41] A. Menchaca-Rocha, F. Huidobro, A. Martinez-Davaloz, K. Michaelian, A. Perez, V. Rodriguez and N. Carjan Journal of Fluid Mechanics 346, 291–318 (1997)

[42] R. Fries, V. Greco, P. Sorensen Annual Review of Nuclear and Particle Science 58, 177-205 (2008)

[43] V. Greco, C.M. Ko and R. Rapp, Physics Letters B 595, 202–208 (2004)

[44] J. S. Goldstein, A. J. Mazzella, Il Nuovo Cimento B 21, 142–150 (1974)

[45] F. Family, P. Meakin, J. M. Deutch, Physical Review Letters Vol. 57, Nr. 6, 727–730 (1986)

[46] P. Carruthers, M. Duong-Van, Physics Letters B 131, 116–120 (1983)

[47] Y. Yoshii, H. Saio, Astrophysical Journal 295, 521–536 (1985)

[48] A.M. Mazzone, Journal of Computer-Aided Materials Design 7, 133–141 (2000)

[49] S. Scherer, H. Stocker, Proceedings of the International Symposium ”Atomic Cluster Collisions”, St. Petersburg, Russia, July 18-21, 2003, 169–180 (2004)

[50] S. Lovejoy, H. Gaonac’h, and D. Schertzer, J. Geophys. Res. 109, BB11203 (2004)

[51] L. Zaichik, V. Alipchenkov, High Temperature 46, 666-674 (2008)

[52] D. Han, X. Sheng Zhang, W. An Zheng, Acta Mathematica Sinica 24, 121-138 (2008)