

Research Paper

Inferring the turbulent breakup of colloidal aggregates using Graph Neural Networks

M. Buzzicotti ^{a,b}, M. Cencini ^{c,b}, G. Cimini ^{a,b}, M. Vanni ^d, A.S. Lanotte ^{e,f} ^{*}

^a Dept. Physics, University of Rome "Tor Vergata", Via della Ricerca Scientifica 1, Rome, 00133, Italy

^b INFN, Sez. Roma 2, Via della Ricerca Scientifica 1, Rome, 00133, Italy

^c Istituto dei Sistemi Complessi, CNR ISC, Via dei Taurini 19, Rome, 00185, Italy

^d Dip. Scienza Applicata e Tecnologia DISAT, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, 10129, Italy

^e Institute of Nanotechnology, CNR NANOTEC, Via Monteroni, Lecce, 73100, Italy

^f INFN, Sez. Lecce, Via Arnesano, Lecce, 73100, Italy

ARTICLE INFO

Keywords:

Aggregates

Fragmentation

Graph Neural Networks

Turbulent flows

Stokesian dynamics

ABSTRACT

Solid aggregates in turbulent suspensions may break under the action of shear stresses. We explore the use of Graph Neural Networks (GNN) to infer aggregate fragmentation once the aggregate structure and flow velocity gradients are known. We consider two models: the first GNN is a classifier, trained to distinguish aggregates that break from those that do not; the second GNN is a regression model, trained to predict the maximal tensile force within each aggregate in a given flow condition. We show that both models complete their task with a high statistical accuracy, also generalizing to aggregates of different sizes, and generally performing better than the statistical prediction based on mean field quantities. This work paves the way for future use of GNN to quantify aggregate breakup in a large population of aggregates suspended in complex flow configurations, as it takes place in the wet production of fine powders and in the transport of sediments in environmental flows.

1. Introduction

In turbulent suspensions, the breakup of agglomerates – material structures formed by an ensemble of fragments or primary particles or monomers compacted together – due to the hydrodynamic stress generated by the surrounding fluid is a key process in a wide range of applications, in both natural and engineering contexts. Examples include rubber compounding (Manas-Zloczower, 2012), the production of micro-particles via wet synthesis (Gavi et al., 2010), the rheological behavior of suspensions (Bilodeau and Bousfield, 1998), and sediment transport in natural flows or wastewater streams (Kaveh and Malcherek, 2025). Despite the broad scope of these applications, only a few experimental studies have analyzed the breakup process at the scale of individual agglomerates (Harshe et al., 2011; Saha et al., 2014, 2016; Qi et al., 2025). Most experimental work focuses instead on the global properties of the population of particles – such as the particle size distribution or mean size –, rather than on the mechanism and evolution of individual breakup events, the reason being that agglomerates breakup is very complex and strongly depends on the aggregate structure and flow-field conditions over time. As a result, such studies do not provide tools for accurate prediction of agglomerate breakup processes.

This limitation can be addressed through numerical simulations that couple fluid dynamics with contact mechanics to model aggregate behavior. Several approaches have been proposed in the literature for this purpose. Many of these (based, e.g., on Discrete Element Methods (DEM) and/or Immersed Boundary methods) for the turbulent breakup of aggregates demand complex implementation and considerable computational resources (Saxena et al., 2025; Chen and Li, 2020; Leonelli et al., 2026). As a consequence, their application soon becomes impractical for large populations of agglomerates when directly coupled with complex flow fields obtained by integrating the three-dimensional Navier–Stokes equations. On the modeling side, there have been attempts to derive scaling laws for the critical stress required to initiate breakup in sheared or turbulent colloidal suspensions (see e.g., (Zaccone et al., 2009)).

An alternative strategy, which we explore here, is to develop surrogate models based on neural networks (NN), capable of determining the breakup condition of an agglomerate after appropriate training. These models are generally computationally efficient and could enable simulations involving very large and heterogeneous aggregate populations, thereby providing a realistic representation of process

* Corresponding author at: Institute of Nanotechnology, CNR NANOTEC, Via Monteroni, Lecce, 73100, Italy.

E-mail address: alessandrasabina.lanotte@cnr.it (A.S. Lanotte).

equipment, thus representing a promising avenue of numerical modeling. Recently, Khalifa and Breuer (2023) trained an artificial NN to quantify collision-induced breakage and agglomeration processes of initially nearly spherical agglomerates, and where the DEM simulations were carried out in vacuum. The trained model was then integrated into an Eulerian–Lagrangian simulation framework to study collision-induced breakup in T-junction laminar and turbulent flows, replacing the physical collision model with the efficient neural surrogate (Khalifa and Breuer, 2024).

In this study, we consider turbulent dilute suspensions, focusing on the role of shear stresses, which we solve exactly (within grid resolution) and that possess a wide distribution with large tails in homogeneous flow conditions (Babler et al., 2015). We present a data-driven approach, based on Graph Neural Networks (GNN), which is meant to make predictions on aggregate breakup without performing simulations to directly compute the forces acting on the aggregate. GNN represent a particularly suitable learning paradigm for predicting aggregate fragmentation, because a colloidal aggregate can be naturally framed as a graph-like object: primary particles interact only through the contact bonds established during aggregation, and the transmission of hydrodynamic stresses throughout the structure depends on this connectivity.

In the last few years, GNN (Scarselli et al., 2009; Hamilton et al., 2017; Bronstein et al., 2017) have become a standard tool for machine learning on graph-structured data, that is, data represented as a set of entities (nodes) with features, and of a set of relationships (links) among them. Initially mainly used to solve node classification and link inference tasks (Zhou et al., 2020), nowadays GNNs have a wide span of applications across domains, including chemistry and material science (Reiser et al., 2022; Thomas et al., 2023; Corso et al., 2024). In fluid dynamics, GNNs have been used for different goals. Since they operate on graphs, which can be converted from any mesh geometry, GNNs can ensure a significantly better resolution for relevant regions within the domain with the same mesh size (see e.g. Barwey et al. (2025)), compared with standard convolutional neural networks. Interestingly, the graph structure allows to study particles dynamics in Stokes suspensions, enabling the construction of surrogate models for the mobility tensor of particles subject to hydrodynamical interactions and external forces (Ma et al., 2022). While GNNs have been used to simulate the complex fluid dynamical problems under a wide range of conditions (Sanchez-Gonzalez et al., 2020; Sergeev et al., 2024; Gao and Jaiman, 2024; Filiatraut et al., 2023; Serhani et al., 2024), including GNN-based LES schemes (Serhani et al., 2024) or multiphase flows in porous media (Vu et al., 2025), their application to solid aggregates in turbulent suspensions remains a relatively unexplored research area.

Here we develop two different GNN models: the first one is simply meant to classify whether a single aggregate will break or not; the second is meant to infer the maximal tensile force acting on one of its internal links. Using the latter model, the inferred maximal force can then be compared to a threshold value for rupture to establish whether the aggregate will break or not. We deal with small agglomerates, whose breakup is induced by the hydrodynamic stresses generated by the turbulent flow only (Kusters et al., 1997; De Bona et al., 2014), rather than by the combination of different processes (flow stresses, inter-agglomerate collisions or impacts with the walls). The GNN is trained with the results of a physical model of the process based on a joint Direct Numerical Simulation and Stokesian dynamics – DEM approach that examines rigid isostatic agglomerates with a fixed fractal dimension of immersed in 3D homogeneous and isotropic turbulence. We mostly focus on a systematic investigation of aggregates of given size (i.e. with a fixed number of monomers, see below), albeit with different geometries. This choice is linked to the objective of the work, which is to demonstrate the applicability of GNNs to the prediction of aggregate breakup in a three-dimensional turbulent suspension, a complex system characterized by both large-scale coherent structures and highly non-Gaussian and intermittent small-scale fluctuations. Then,

we extend our analysis to an heterogeneous population of aggregates, to test the generalization ability of our model. Extension to turbulent suspensions in confined volumes will be evaluated in future work based on the results obtained here.

The paper is organized as follows. In Section 2, we introduce the system under consideration and the numerical modeling of the turbulent suspension. Then, in Section 3, we details the databases (DBs) used to train and test the GNNs. The Graph Neural Network models – classifier and regression – are introduced in Section 4, while a statistical test to be compared to the Neural Networks results is presented in Section 5. Section 6 is devoted to discussing the performances of the two GNNs on different datasets for homogeneous in size suspensions, the origin of model failures and the generalization ability when considering heterogeneous in size suspensions. We summarize our work in Section 7.

2. Numerical modeling of turbulent suspensions

The simplest numerical methods to study aggregate breakup use the free-draining approximation, where each primary particle is assumed to be unaffected by the flow perturbations caused by neighboring particles (Becker et al., 2009; Eggersdorfer et al., 2010; Schutte et al., 2018; Ruan et al., 2020). Estimates of the screening effect of hydrodynamic forces within agglomerates have been based on the fraction of the primary particle surface area directly exposed to the external fluid flow (Higashitani et al., 2001; Fanelli et al., 2006). More sophisticated semi-empirical methods, that relate screening to the solid volume fraction or aggregate geometry, were later adopted by Yao and Capecelatro (2021) and Yu et al. (2023). These approaches fall within the Eulerian–Lagrangian framework, in which aggregates are treated as point-like entities.

A more computationally demanding alternative involves full resolution of the fluid flow at the scale of the primary particles. This route was taken by different authors by means of Lattice-Boltzmann simulations to investigate the restructuring and breakup of agglomerates in shear (Saxena et al., 2022, 2023), accelerated (Saxena et al., 2025) and turbulent flows (Derksen, 2013).

For colloidal aggregates with very small particle Reynolds number, $Re_p = \gamma_{eff} R^2 / \nu \ll 1$ (R being the aggregate size in terms of its radius of gyration, ν the fluid kinematic viscosity and γ_{eff} the effective shear rate), Stokesian Dynamics (SD) offers an equally accurate alternative, enabling precise computation of the hydrodynamic forces acting on the aggregates. Compared to techniques that fully resolve the solid–liquid interfaces, and the hydrodynamic forces/torques acting on each primary particle, SD is less computationally expensive and allows for the study of systems with several hundreds of primary particles, in contrast to full CFD simulations, which typically handle only a few dozen of them. Over the years, Stokesian dynamics has been used to investigate the breakup and restructuring of soft and rigid agglomerates in uniform shear flows (Harada et al., 2006; Seto et al., 2011; Vanni and Gastaldi, 2011; Harshe and Lattuada, 2012; Frungieri and Vanni, 2021), elongational flows (Ren et al., 2015), complex confined flows (Vasquez Giuliano et al., 2022; Frungieri et al., 2022), and isotropic homogeneous turbulence (De Bona et al., 2014). For aggregates that remain rigid, without deforming, until the onset of breakup, the Stokesian dynamics simulation approach can be significantly accelerated using the procedure proposed by Vanni (2015).

We consider a colloidal suspension of small aggregates whose center of mass, $\mathbf{X}_{cm}(t)$, evolves as neutrally buoyant tracer particle

$$\frac{d\mathbf{X}_{cm}}{dt} = \mathbf{V}_{cm}(t), \quad (1)$$

$\mathbf{V}_{cm}(t) = \mathbf{v}(\mathbf{X}_{cm}(t), t)$ being the fluid velocity at the aggregate center of mass. Eq. (1) is valid for aggregates of size R much smaller than the Kolmogorov scale of the turbulent flow, η , and whose density matches that of the fluid. Moreover, we consider a very dilute suspension, where

Table 1

Parameters of DNS in simulation units (su). Microscale Reynolds number Re_λ corresponding to a resolution $N^3 = 256^3$ grid points; root-mean-square velocity u_{rms} ; the mean kinetic energy input rate ϵ_{in} equal to the mean kinetic energy dissipation rate ϵ_v ; root-mean-square velocity gradient $\gamma_{eff} = \sqrt{\epsilon_v/\nu}$; kinematic viscosity ν ; Kolmogorov length-scale $\eta = (\nu^3/\epsilon_v)^{1/4}$; integral scale L ; Eulerian large-scale eddy turnover time $T_E = L/u_{rms}$; Kolmogorov timescale $\tau_\eta = \sqrt{\nu/\epsilon_v}$; sampling time along Lagrangian trajectories dt_{Lag} ; total integration time T_{tot} ; total number of advected Lagrangian tracers N_p .

Re_λ	u_{rms}	ϵ_{in}	γ_{eff}	ν	η	L	T_E	τ_η	dt_{Lag}	T_{tot}	N_p
90	3.3	2.3	15.1	0.01	0.024	π	0.95	0.07	0.006	36	524 000

aggregate collisions, hydrodynamic coupling and particle feedback onto the flow can be neglected (Balachandar and Eaton, 2010).

The fluid phase evolves according to the three-dimensional incompressible ($\nabla \cdot \mathbf{v} = 0$) Navier–Stokes equations,

$$\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{1}{\rho_f} \nabla p + \nu \nabla^2 \mathbf{v} + \mathbf{f}, \quad (2)$$

where p is the pressure field, ρ_f is the fluid density and $\mathbf{f}$ is an external three-dimensional forcing. At statistical steady state, the power input, $\epsilon_{in} = \langle \mathbf{f} \cdot \mathbf{v} \rangle$, balances kinetic energy dissipation, $\epsilon_v = \nu \langle |\nabla \mathbf{v}|^2 \rangle$, where angular brackets stand for spatial average over the fluid volume.

Direct Numerical Simulations (DNS) of the Navier–Stokes equations are solved adopting a standard pseudo-spectral approach, fully dealiased with the two-thirds rule, within a cubic, periodic domain of length 2π . The kinematic viscosity is such that $\eta \simeq \Delta x$, i.e. the Kolmogorov length-scale η is of the order of the grid spacing Δx . This implies that turbulent dissipative range dynamics is well resolved. The statistically homogeneous and isotropic random forcing $f(\mathbf{x}, t)$, continuously supplying kinetic energy to the flow, acts at large scales, around the wave-number $k_f = 1$, and it is the solution of a second-order Ornstein–Uhlenbeck process (Sawford, 1991). In our numerical set-up, there is no imposed mean shear, and the effective shear rate is estimated as $\gamma_{eff} = \sqrt{\epsilon_v/\nu}$ (see later on). Relevant parameters of the DNS are summarized in Table 1 and are given in simulation units.

Aggregates moving in the turbulent flow are subject to internal stresses due to the interaction of the aggregates with the local flow field. In the colloidal suspension here considered, since $Re_p \ll 1$, significant variations in the fluid velocity gradients take place on distances much larger than the size R of the aggregates: this means that the particle size is assumed to be much smaller than the Kolmogorov scale of the flow, $R \ll \eta$. Consequently, in evaluating the hydrodynamic stresses, it is reasonable to neglect the curvature of the velocity profiles and assume that the aggregates are surrounded by a linear flow field with uniform velocity gradient. Under this assumption, Stokesian dynamics (Brady and Bossis, 1988) can be effectively adopted to numerically determine the hydrodynamic forces, torques and stresses acting on each aggregate. The method is mesh-less and provides the relationship between the hydrodynamic interactions acting on the particles and their velocity. The velocity $\mathbf{V}_i$ at each of the i -th primary particle position $\mathbf{x}_i$ is given by

$$\mathbf{V}_i = \mathbf{V}_{cm}^\infty + \mathbf{\Gamma}^\infty \cdot (\mathbf{x}_i - \mathbf{X}_{cm}), \quad (3)$$

where $\mathbf{V}_{cm}^\infty$ is the undisturbed flow velocity and $\mathbf{\Gamma}^\infty$ is the velocity gradient tensor, $\Gamma_{ij} = \partial_j v_i$, evaluated at the aggregate center of mass. Additionally, the aggregates are assumed to be rigid, and hydrodynamic forces – associated to both translational and rotational motions – are redistributed as internal stresses over the aggregate (Vanni and Gastaldi, 2011; Vanni, 2015). SD simulations allow to evaluate the stress distribution and the location where the stress is highest, i.e. the bond for the onset of breakup.

The results presented here are obtained by first solving the aggregates motion as point-like tracers evolving in the three-dimensional turbulent flow, and computing the instantaneous value of the velocity gradient matrix at their position. The SD dynamics is then integrated offline to compute the maximal tensile force. Further details can be found in De Bona et al. (2014).

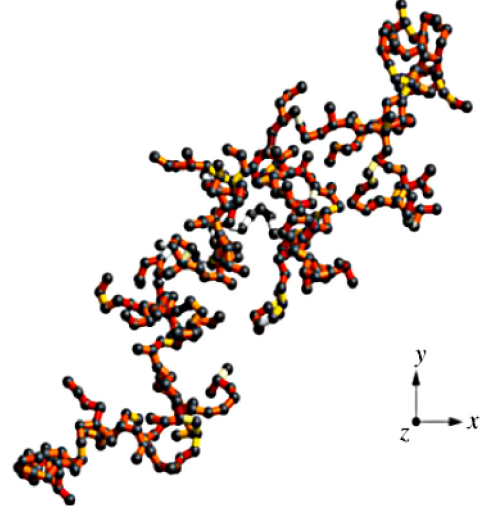


Fig. 1. Structure of a cluster–cluster aggregate of fractal dimension $D_F = 1.9$, made up on $N_M = 384$ primary particles of radius a , and $N_L = N_M - 1$ links among them. The gyration radius of the aggregate scales with the primary particle radius and the number of primary particles as $R \simeq a N_M^{1/D_F}$.

We consider colloidal suspensions of rigid and isostatic aggregates made up by few hundreds of primary particles of radius a . They are generated numerically by using a tunable cluster–cluster method (Lattuada et al., 2003), capable of producing aggregates which do not have spherical symmetry and obey the following relation

$$N_M \simeq \left(\frac{R}{a} \right)^{D_F} \quad (4)$$

where N_M is the number of primary particles composing the aggregate, and R is the gyration radius of the aggregate. For a systematic study, we set $D_F = 1.9$ and $N_M = 384$, and the typical shape is given in Fig. 1, then we will briefly consider aggregates with different values of N_M at the end of our work. Agglomerates with such features are quite common, as they are typically obtained by fast coagulation processes of sub-micron primary particles. The value of fractal dimension is a measure of the mechanism of aggregation, with smaller values being associated with more open structures. The condition of isostaticity implies that there are no closed loops of particles in the structure and all the chains of primary particles in the aggregate are open on one side, so that the failure of a single bond results in the breakup of the aggregate into two fragments.

3. Aggregates statistics and ground-truth data

When aggregates move and rotate in the turbulent flow, the hydrodynamics stresses produce a turnover between compression and elongation forces acting on each inter-particle link. For isostatic aggregates, breakup occurs whenever the maximal internal tensile force (i.e. on the most loaded bond) exceeds the critical pull-off value, given by contact mechanics (Johnson, 1985), $F_{cr} \propto a\sigma$, where σ is the surface energy at the contact. Hence, to determine the occurrence of breakup the key observable is the maximal force F_{max} to be compared to the

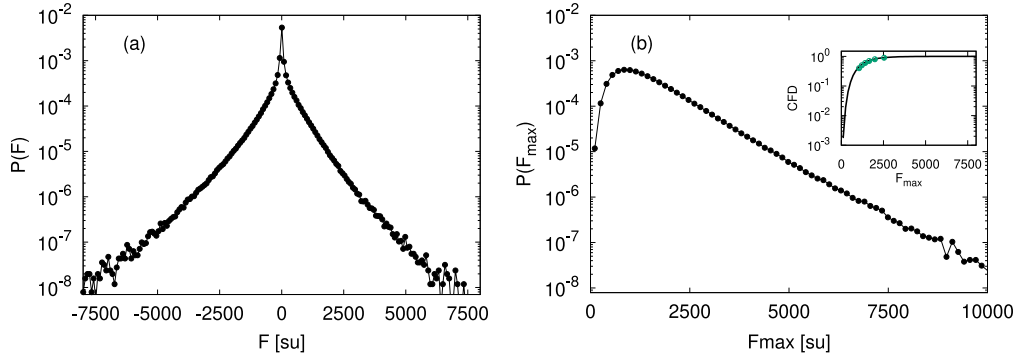


Fig. 2. (a) PDF of the tensile force F , exhibiting tails with exponential behavior (De Bona et al., 2014). (b) PDF of the maximal tensile force $F_{\max}$; in the inset, the cumulative distribution of $F_{\max}$ (continuous line). The green circles plotted on top of the CDF identify the six threshold values for the maximal tensile force, $F_{th1} = 1052$; $F_{th2} = 1220$; $F_{th3} = 1417$; $F_{th4} = 1653$; $F_{th5} = 1988$; $F_{th6} = 2539$, corresponding to the CDF being equal to 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9, respectively.

pull-off threshold value. Since breakup occurs upon the failure of a single bond, aggregates with multiple bonds with local tensile force above the critical threshold are considered nonphysical.

In Fig. 2(a), we plot the probability density function (PDF) of the normal tensile force averaged over all the aggregates in the suspension, that exhibits a stretched exponential behavior and is symmetrical. We note that at the level of the single bond, this PDF may be skewed towards negative values (if compression prevails) or positive values (if extension prevails). Fig. 2(b) shows the PDF of the maximal tensile force, $F_{\max}$, while the inset displays its cumulative distribution function (CDF). From the CDF, six threshold values for the breakup are defined, corresponding to the CDF ranging, with equal spacing, from 40% to 90% (as detailed in the figure caption). These thresholds define six distinct aggregate families, such that for each of them breakup occurs when the tensile force on the most loaded bond, $F_{\max}$ exceeds the threshold value, i.e., $F_{\max} > F_{th}$.

For our purpose of developing a data-driven tool, it is important to construct unbiased datasets, containing random, independent realizations. This helps to avoid temporal correlations that could possibly influence the NN learning stage. From DNS data, we selected 2×10^6 velocity gradient realizations, that are statistically independent. Additionally, we considered 1000 different aggregate shapes, all characterized by the same fractal dimension D_F and number of monomers N_M , but differing in shapes. Each aggregate-shape was then randomly rotated through 2000 different orientations, and each configuration was exposed to a different velocity gradient. Finally, for each of these realizations, we computed the Stokesian dynamics, and obtained the tensile force on each aggregate bond from which we extract the value for the most loaded bond, $F_{\max}$. These data have been further split into different datasets for the learning and testing stages, summarized in Table 2. Details of the dataset preparation can be found in Appendix A.

3.1. Database description: Classification task

To classify broken and unbroken aggregates, we selected realizations for which $F_{\max}$ exceeds a threshold value on a single bond only, in order to respect the isostatic assumption of the aggregates: by doing so, we generated a database containing 720.000 examples out of which 576.000 were used as learning dataset. These are further organized into six families, corresponding to the different rupture thresholds. In other words, each (single-rupture) aggregate has been selected by measuring the maximal tensile force over all the internal bonds and comparing it to one randomly selected threshold value from the six selected values.

As explained above, after having generated a large DB of 2×10^6 aggregates, we impose the threshold for the rupture and eliminate all aggregates that possess more than one bond with maximal tensile

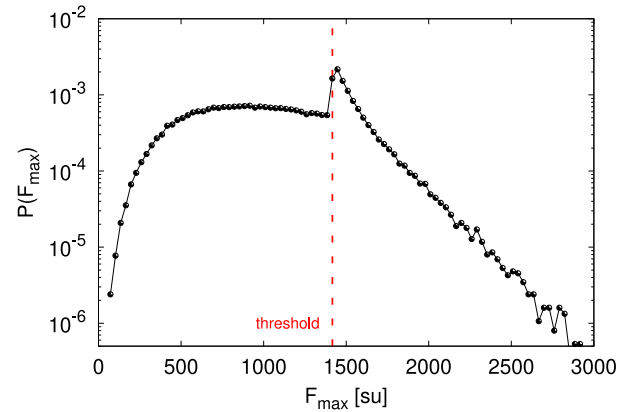


Fig. 3. PDF of the maximal tensile force $F_{\max}$ for the aggregates that either do not break, or break for a single bond failing when $F_{\max} \geq F_{\text{threshold}}$, with $F_{th3} = 1417$ (vertical red dashed line).

force above threshold. This procedure is physically justified by the fact in a physical experiment, as soon as the maximal force reaches the threshold value the aggregate breaks and that it is very unlikely that two bonds overcome the threshold value at the same instant. As a consequence, as we do not follow the evolution of aggregates but generate them from random rotations and uncorrelated instances of the velocity gradients, those samples that present more than one rupture should be considered as not physically realizable, as they are expected to be already broken. From a statistical point of view this procedure causes a conditioning that modifies the statistics of the maximal tensile force (Fig. 2(b)). This effect is exemplified in Fig. 3. Most of the broken aggregates experience a value of $F_{\max}$ close to the threshold value, and the resulting PDF decreases very sharply as the value of $F_{\max}$ increases above the threshold value. Hence, for broken aggregates, the maximal tensile force is mostly concentrated just above the pull-off value for the bond failure. Differently, the distribution of the unbroken aggregates is broader, since for these aggregates any force smaller than the threshold ensures the absence of a link failure. However, the probability close to the threshold is modified by the threshold value also for the unbroken aggregates and this impacts the performance of the GNN classifier, as we will see later in the result section. This is generally true for all thresholds, and it is related to the single-bond rupture mechanism.

To mitigate the statistical prevalence of broken cases very close to the threshold (which are the most challenging to predict), and reduce the statistical imbalance of broken versus unbroken cases close the threshold values, we opted for the following choice: the learning database contains 33% occurrences of broken aggregates and 67% occurrences of unbroken aggregates for each of the six thresholds. After

the learning or training stage, the GNN classifier performances were tested on three different DBs, that we briefly describe in the following (see also Table 2).

Cl-imb: A test-set was generated using velocity gradients and random rotation of the aggregates that were never seen during training. This test-set is unbalanced, consistently with the training DB, and contains approximately 67% unbroken aggregates and 33% broken aggregates.

Cl-bal: A balanced test-set containing 50% broken, 50% unbroken aggregates. It is obtained by randomly selecting a subset of realizations of the previous dataset, Cl-imb.

Cl-rnd: A balanced test-set (50% broken, 50% unbroken), made of realizations exhibiting a single rupture with respect to a randomly selected threshold value for the maximal tensile force. In particular, for each sample, a random threshold value is drawn with uniform probability in the range [5%; 95%] of the CDF of the maximal tensile force (corresponding to the interval $F_{\max} \in [432; 3100]$). After this random selection, the threshold is compared to the maximal tensile force computed by Stokesian Dynamics, with the constraint that no more than a link has tensile force above threshold, to determine if it is a broken or unbroken aggregate realization. The purpose of this DB is to test the classifier model ability to generalize with respect to threshold values never seen during the training.

3.2. Databases description: Regression task

To train and test the GNN regression model aimed to infer the maximal tensile force for each aggregate in a given flow realization, we prepared another dataset. In the regression model, the rupture threshold is not known and there is no *a priori* distinction between broken and unbroken aggregates. The learning database contains 525.000 samples obtained from a collection of about one thousand different aggregate shapes, exposed to different rotations and fluid velocity gradients; the GNN regression model was then tested on the following two DBs.

Re-test: A test-set with about 132.000 samples of aggregates exposed to velocity gradient and rotation realizations, never seen during the training. The geometrical shapes of the aggregates composing this test database have been seen by the model during the training stage.

Re-gen: A test-set with about 128.000 samples of aggregates whose shape, velocity gradient and rotation realizations had never been seen during the training stage. Hence, this dataset provides a basis for evaluating the extent to which the GNN can generalize its performance to unseen geometries of the aggregates. The new shapes still have the same fractal dimension, D_F , and the same number of primary particles, N_M , as the aggregates used for training.

4. Graph neural network models for aggregate breakup

The basic architecture of the two models, namely the *classifier* and the *regression* GNNs, is sketched in Fig. 4: the models share several general features and most architectural implementation details. In particular, the input data consist of an aggregate represented as a graph, where each monomer corresponds to a node and the connections are encoded in a bidirectional adjacency matrix, whose elements denote physical contact between particles. For each i th node $i = 1, \dots, N_M$, the input features form a six-dimensional vector

$$(x_i, y_i, z_i, V_{x,i}, V_{y,i}, V_{z,i}), \quad (5)$$

with the node's position coordinates relative to the aggregate's center of mass $\mathbf{X}_{cm}$, and the velocity at each primary particle position. The velocity at each node is given by Eq. (3).

Differently from the regression model, the classifier additionally receives the threshold value F_{th} corresponding to pull-off force. The threshold information is encoded through a fully connected layer that takes the one-dimensional real value as input and maps it into a 64-dimensional latent vector. This information is essential for predicting

the rupture, whereas it is not required for estimating the maximum tensile force experienced by the aggregate's links. In the latter case, the likelihood of breakup can be evaluated *a posteriori* by comparing the inferred maximal tensile force with the threshold value.

In the training stage, the classifier learns to distinguish between broken and unbroken aggregates with respect to six different threshold values of the maximal tensile force, as discussed above. The input graph, in both models, is processed through four successive concatenations of graph convolutional layers. Each convolution is performed according to the GraphSAGE scheme (Hamilton et al., 2017), with a max aggregation operation applied after each convolution. Specifically, the node embeddings are updated at the m th layer as

$$\mathbf{h}_i^{(m)} = \mathbf{W}_1^{(m)} \cdot \mathbf{h}_i^{(m-1)} + \mathbf{W}_2^{(m)} \cdot \max \left\{ \mathbf{h}_j^{(m-1)} : j \in \mathcal{N}(i) \right\}, \quad (6)$$

where $\mathbf{h}_i^{(m)}$ denotes the embedding of node i at layer m , $\mathcal{N}(i)$ is the set of neighbors of node i (the particles connected to i by cohesive bonds), $\max$ indicates element-wise maximum aggregation. $\mathbf{W}_1^{(m)}, \mathbf{W}_2^{(m)}$ are two learnable weight matrices that project the 6-dimensional input feature vector (5) into a 64-dimensional embedding space. In our framework, we do not apply a non-linear activation function after each layer.

At the end of the four graph convolutions, the output at each node is constructed by concatenating the four 64-dimensional feature vectors obtained from each layer, resulting in a final 256-dimensional representation for each node. At the final stage of the GNN convolutions, a global max pooling operation is applied over all node embeddings to obtain a single 256-dimensional latent vector representing the entire input graph. This pooling operation is independent of the number of nodes in the input graph enabling, in principle, the architecture to process aggregates of varying sizes.

In the classifier, the 64-dimensional embedding, encoding the threshold information, is concatenated at this point with the 256-dimensional output of the graph embedding, resulting in a combined 320-dimensional latent vector. At this stage, the input embeddings for both models are mapped onto the desired one-dimensional output through six fully connected layers. These layers gradually reduce the input dimension while applying a nonlinear activation function. The first five dense layers use a ReLU activation, while the activation of the final output layer is model-specific, in particular: the classifier employs a softmax activation, while the regression model uses a linear activation.

The *classifier* GNN outputs a rupture probability in $[0, 1]$, with 1 denoting breakage. Training consists in minimizing the binary cross-entropy loss, L_{CE} , between the model rupture probability and the ground-truth label Y^{true} :

$$L_{CE}(p, Y^{true}) = -\frac{1}{N} \sum_{k=1}^N [Y_k^{true} \log(p_k) - (1 - Y_k^{true}) \log(1 - p_k)], \quad (7)$$

where N is the batch size in the learning process.

The *regression model* is trained to predict the maximal tensile force exerted by the fluid on the aggregate, outputting a positive real number. In this case, during the training the normalized mean squared error (MSE) loss, L_{MSE} , is minimized:

$$L_{MSE} = \frac{1/N \sum_{k=1}^N [(F_k^{pred} - F_k^{true})^2]}{\langle (F^{true})^2 \rangle_{DB}} = \frac{\langle (F^{pred} - F^{true})^2 \rangle}{\langle (F^{true})^2 \rangle_{DB}}, \quad (8)$$

where N is the batch size in the learning process, and the average $\langle \bullet \rangle_{DB}$ is over the whole learning database.

Fig. 5 illustrates the evolution of accuracy for the classifier (left) and mean squared error, Eq. (8), for the regression model (right), over training epochs for the training and test datasets. The curves obtained during GNN testing still vary across epochs because test loss and accuracy are repeatedly computed during training after each update to the network's parameters. In this workflow, the model is optimized exclusively with respect to the loss evaluated on the training data, while the test datasets are used to periodically assess performance.

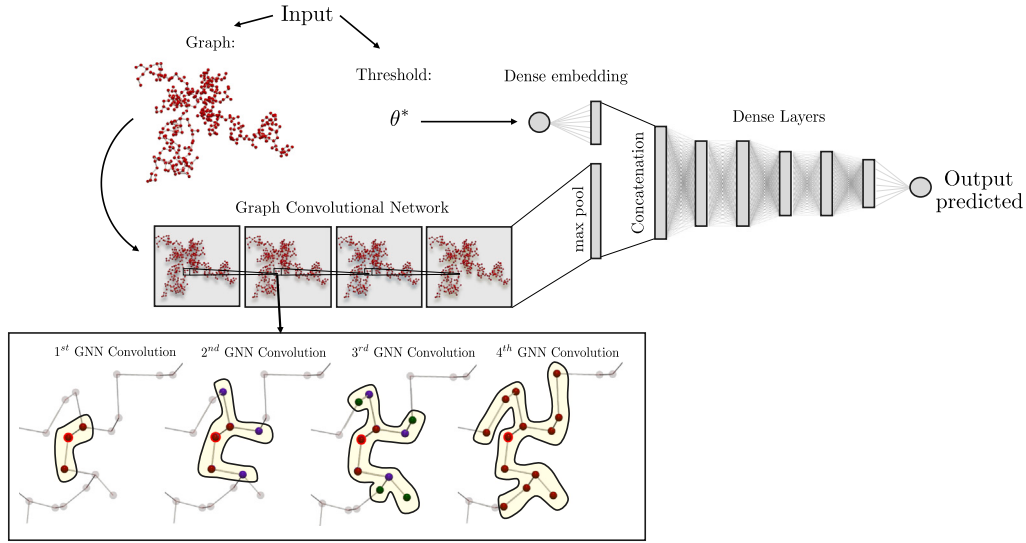


Fig. 4. A scheme of the end-to-end prediction task with the GNN classifier producing a binary output; for the regression model, the scheme is very similar, but there is no threshold value given in input, and the output is a positive real number.

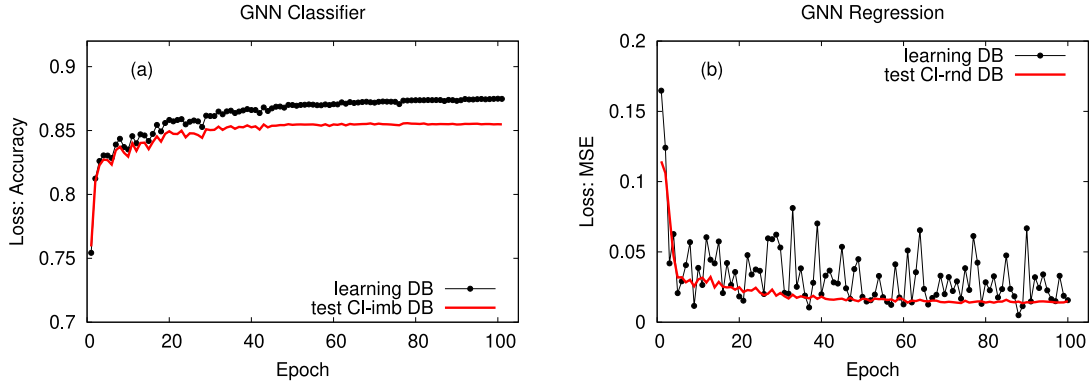


Fig. 5. (a) The evolution of success probability (or accuracy) with the number of epochs for the GNN classifier model on the learning dataset (filled circles) and the test CI-imb dataset (continuous red line). (b) The evolution of loss function, L_{MSE} , with the number of epochs for the GNN regression model on the learning dataset (filled circles) and the test CI-rnd dataset (continuous purple line).

Monitoring both the training and test metrics is essential to verify that the learning process is converging. A decrease in training loss alone is indeed insufficient, because the model may improve on the training set due to overfitting, resulting in degraded generalization. By comparing the two training and testing curves, one can identify when the model stabilizes and determine if the performance on unseen data reflects the improvements observed during training.

5. Statistical test for aggregate break-up

In order to benchmark the performances of the GNN models for the classification of aggregate breakup, we devised a simple statistical model for detecting rupture, which is illustrated in the following.

In our set-up, the breakup of aggregates is due to hydrodynamic stresses acting on the links between the primary particles composing the aggregate (Kusters et al., 1997; De Bona et al., 2014). In simple shear flows, the hydrodynamic stress is given by the uniform shear rate, γ . In turbulent flows, it can be quantified in terms of the effective shear rate

$$\gamma_{eff}(\mathbf{x}, t) \equiv \sqrt{S_{ij}S_{ij}}, \quad (9)$$

where summation over i, j is implied and S_{ij} denotes the symmetric part of the velocity gradient tensor, i.e. $S_{ij} = (\Gamma_{ij} + \Gamma_{ji})$ and $\Gamma_{ij} = \partial_j v_i$. As it is well known, we can write $\gamma_{eff} = \sqrt{\epsilon_v}/\nu$, where $\epsilon_v(\mathbf{x}, t)$ is the

local kinetic energy dissipation that, in turbulence, is known to exhibit strong fluctuations both in space and time (Frisch, 1995; Meneveau and Sreenivasan, 1991; Benzi et al., 2009).

For brittle isostatic aggregates, as considered here, we assume that breakup takes place as soon as γ_{eff} is larger than a given threshold γ_{cr} . Actually, it has been shown that turbulent breakup may not only depend on the instantaneous value of γ_{eff} , but also on the instantaneous orientation of the aggregate with respect to the flow field (De Bona et al., 2014). For simplicity, we ignore such an effect.

We computed the effective shear rate γ_{eff} conditioned on the value of F_{max} , $\langle \gamma_{eff} | F_{max} \rangle$, in the same dataset used to train the GNN-CI model. The result is plotted in Fig. 6. We observe that such a quantity behaves in very good approximation as a power law,

$$\langle \gamma_{eff} | F_{max} \rangle \propto c F_{max}^b, \quad (10)$$

with the exponent $b \approx 2/3$, and c a prefactor. We note that a dimensional relation between the tensile force and the effective shear rate is expected for sub-Kolmogorov aggregates, namely $F \propto \mu R_g^2 \gamma_{eff}$, where μ is the dynamic viscosity of the flow. However, this dimensional argument does not constrain how the maximal force value depends on the effective shear rate.

While further investigation is needed to better assess the observed power law, we recall that relationships of this kind are known in turbulent aggregates, e.g. the typical size of an aggregate is known to be

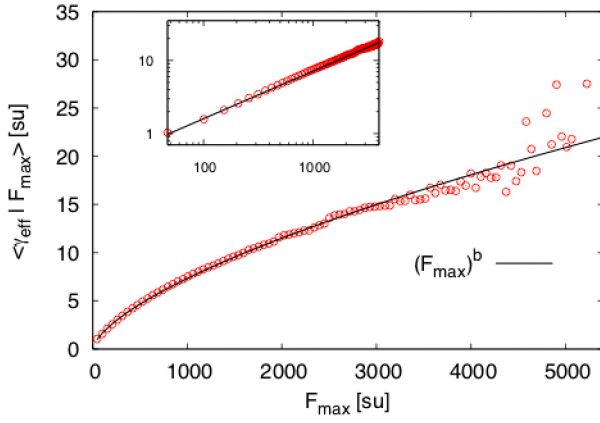


Fig. 6. Conditional effective shear rate $\langle \gamma_{eff} | F_{max} \rangle$ vs F_{max} computed on the learning dataset of the GNN classifier (576.000 samples). The solid curve shows the fitted behavior, Eq. (10). In the inset, the same plot in log-log scale, where a linear fit is applied in the range $F_{max} \in [20 : 3000]$. The fit gives the exponent $b = 0.65 \pm 0.02$ and the prefactor $c = 0.08 \pm 0.01$.

proportional to an inverse power of the average effective shear strength (see, e.g., the review of Jarvis et al. (2005) and references therein).

The observed power law of Eq. (10) can be used to establish a statistical test for the breakup of an aggregate, provided one knows, γ_{eff} , and the threshold force, F_{th} . These inputs are the same that the GNN classification model receives, since the effective shear rate γ_{eff} is implicitly available through the flow velocities at the positions of the primary particles. In particular, it is natural to assume that the aggregate breaks when $\gamma_{eff} > cF_{th}^b$, while it remains unbroken otherwise. Applying the above statistical test on the test-set Cl-imb, we find that this test works correctly for $\approx 74.46\%$ of the samples, while in Cl-rnd 69.69%. The difference between the performances in the two DB may be due to the fact that the statistical test and in particular the constant A in Eq. (10) was fitted on the data with 6 thresholds and so there may be some tiny statistical differences from the random threshold DB.

6. Results and discussion

In this section we will first present the results of the two models in classifying broken and unbroken aggregates in the test databases, then we will discuss the origin of the errors in classification specific of the two models and finally we will details the performances of the regressor model in predicting the maximal tensile force.

6.1. Classification task

As discussed in the previous section, the classifier model GNN-Cl directly outputs the rupture probability. In particular, we remark that the GNN-Cl makes a correct prediction when the output is either 1.AND. $F_{max} \geq F_{th}$ or 0.AND. $F_{max} < F_{th}$. In order to use the GNN regression model for classifying the aggregates, we proceed in two steps. First, the regression model outputs the value of the maximal tensile force, for each sample. Then, the predicted maximal force is compared with the assigned threshold for the specific realization to establish whether the predicted value is above (broken) or below (unbroken) threshold. The discussion on the regression ability, i.e. the performances of the GNN-Re in predicting the correct F_{max} , will be postponed to the next subsection, here we focus on the classification task.

Table 2 summarizes all results, together with information about the models, datasets and loss functions used during the training and test stages. In particular, the success probability is defined as the number of correctly classified samples normalized to the total number of samples.

We see that the GNN-Cl works very well, correctly classifying at best 85.5% of the samples, while failing in the 14.5% of the cases. Also in all datasets, the GNN-Cl works better than the GNN-Re, which at best correctly classifies 80% of the samples. Finally, we can see that the statistical test of Section 5 has poorer results compared to both the GNN models.

Besides the success probability, Table 2 presents the so-called F_1 -score or F-measure, giving the predictive performances for the classification of broken aggregates (Fawcett, 2006). The F_1 -score is the harmonic mean of precision and recall,

$$F1 = 2 \frac{p \times r}{p + r} \quad (11)$$

where the precision (p) is the ratio between the number of aggregates correctly classified as broken (true positive) and all aggregate that were predicted to be broken (all positive), and the recall (r) is the ratio between the number of aggregates correctly classified as broken (true positive) and those which are actually broken (true positive plus false negative). A high value of the F_1 score means that the classifier is both precise and sensitive: results show that the GNN-Cl outperforms the classification results of both the regression and the statistical models. Moreover, we observe that for both the DB CL-imb and CL-bal, the success probability and the F_1 values remain high. This confirms that having trained the GNN-Cl with an unbalanced dataset does not impact model performances: when applied to balanced test dataset, the GNN-Cl still performs very well.

To disentangle instances where the GNNs fail, we look at wrong results as follows: we define *false positive* the unbroken aggregates that the GNN classifies as broken (UCB), and *false negatives* broken aggregates that the GNN classifies as unbroken (BCU). These data are also reported in Table 2. We can see that in general the regressor model has a larger probability to detect false negatives, while the classifier is more balanced in Cl-imb and Cl-bal while it also produces more false negative in the DB Cl-rnd. Also the statistical test tends to generate more false negative than positive.

As shown in Fig. 3, once a threshold is assigned and the condition to have at maximum one bond with the tensile force above threshold is imposed, the PDF of the maximal force strongly deviates from the unconditional PDF, shown in Fig. 2(a). In particular, the form of the PDF is strongly modified close to the threshold and for values above it. It is thus natural to expect that the distance of the maximal tensile force from the threshold plays an important role in the classification task. To better inspect this, we can look at the success probability conditioned on the relative distance between the maximal tensile force of each aggregate and the threshold force associated to the breakup condition, $X = \frac{F_{max} - F_{th}}{F_{th}}$. In Fig. 7, we show the conditional success probability of the two GNN models and for the statistical model, for the datasets Cl-imb and Cl-rnd in the left and right panel, respectively. For each model, datapoints that lie above the dataset average success rate (dashed color lines) refer to samples for which the model predictions work better than its average. As expected, correct predictions are more and more frequent as the distance from the threshold value is larger and larger: in other words, when the true force is much larger (or smaller) than the assigned threshold, the model correctly predicts the rupture (or its absence, respectively).

The region of force values close to the threshold is the one where making correct predictions is the most difficult, and all models concentrate their failures, but not in the same way.

First, we note that GNN-Cl is better than GNN-Re at classifying in particular in the region above threshold, i.e. for broken aggregates. In this region, the statistical test is the one performing the worst. The GNN-Re model, as detailed in the next subsection, tends to underestimate F_{max} when its true value is large, and this is the main source of false negatives. However, we should take into consideration two facts: first, regression models have been trained and optimized to predict F_{max} and not to classify; second, regression models have been trained

Table 2

Col 1: name of the test database; col 2: Model name: Cl and Re correspond to the GNN classifier and regression models, respectively, while SM corresponds to the statistical model discussed in Section 5; col 3: Loss function based on binary cross-entropy (CE) or mean-squared error (MSE); col 4: total number of aggregate samples; col 5: number of broken aggregates; col 6 number of unbroken aggregates; col 7: total number of correct predictions; col 8: total number of wrong predictions; col 9: false negatives, or BCU, i.e. true Broken aggregates Classified as Unbroken; col 10: false positives or UCB, true Unbroken Classified as Broken; col 11: (unconditional) success probability, giving the correct prediction over the total number of samples; col 12: F_1 -score, giving the harmonic mean of precision and recall. Note the different composition of the two databases: Cl-imb contains about 2/3 of unbroken aggregates and 1/3 of broken ones; differently, Cl-bal and Cl-rnd are both equally split.

DB	Model	Loss	Tot	Broken	Unbroken	Correct	Wrong	BCU	UCB	%success	F_1 -score
Cl-imb	Cl	CE	144 000	47 789	96 211	123 143	20 857	8043	12 814	85.5	0.79
Cl-imb	Re	MSE	144 000	47 789	96 211	115 295	28 705	21 925	6780	80.0	0.64
Cl-imb	SM	–	144 000	47 789	96 211	107 218	36 647	23 257	13 390	74.5	0.57
Cl-bal	Cl	CE	90 000	45 000	45 000	76 454	13 546	7562	5984	84.9	0.84
Cl-bal	Re	MSE	90 000	45 000	45 000	66 282	23 718	20 621	3097	73.6	0.67
Cl-rnd	Cl	CE	200 000	100 000	100 000	166 604	33 396	22 441	10 955	83.3	0.82
Cl-rnd	Re	MSE	200 000	100 000	100 000	150 082	49 918	44 879	5039	75.0	0.68
Cl-rnd	SM	–	200 000	100 000	100 000	139 371	60 629	48 986	11 643	69.7	0.62

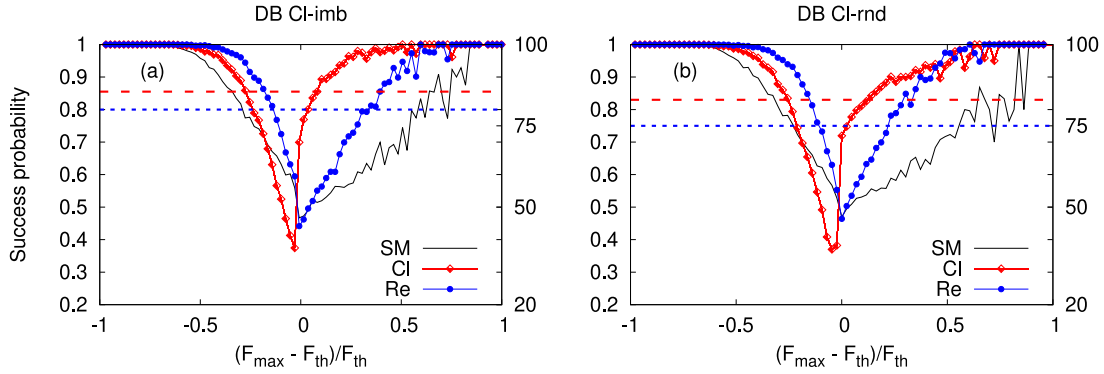


Fig. 7. Success probability (in classification) as a function of the signed relative difference, $(F_{\max} - F_{th})/F_{th}$, between the maximal tensile force $F_{\max}$ and the threshold value, F_{th} . (a) Models are tested on dataset Cl-imb; (b) the same but for dataset Cl-rnd. In each plot, the red curve with diamonds refers to the GNN Classifier (Cl), the blue curve with filled circles refers to the GNN Regression (Re) model, and the continuous line refers to the statistical model. The color horizontal dashed lines correspond to the global success probability on each model (col 11 in Table 2), colors are the same of the model lines.

on an aggregate dataset in which no threshold value was imposed, while classifier models have been trained on samples such that there is only one link with tensile force larger than a given threshold. This implies that the distribution of the forces in the learning databases of the GNN-Cl and GNN-Re models differ, and in particular the training dataset used for the GNN-Re may contain aggregates with multiple links having a tensile force above some given threshold.

In Fig. 7, we have seen that the success probability conditioned on the relative distance from the threshold is not symmetric. Indeed, for the GNN-classifier most of the failures come from samples for which $F_{\max} < F_{th}$, i.e. unbroken aggregates, that have been wrongly classified as broken. These cases, defined (UCB), are on the left part of the distribution, and their success probability is much smaller than the reference value. To better understand the behavior of the classifier we have studied separately the 6 families identified by their thresholds. In Fig. 8, we show the distribution of the relative distance between the true maximal force and the threshold, $(F_{\max} - F_{th})/F_{th}$, for three of the six considered threshold values (the smallest, a central one, the largest). We clearly observe that in the negative region, i.e. when $F_{\max} < F_{th}$, the smaller the threshold the larger is the number of realizations with $F_{\max} \approx F_{th}$, while in the positive region (of broken aggregates) the distribution is independent of the threshold value. This suggests that the greater number of wrong predictions is expected for small thresholds, while it should be less probable to misclassify, for larger and larger thresholds. In other words, the performances of the GNN classifier are hence expected to depend on the threshold value, as it is confirmed by the results for specific thresholds reported in Table 3. As one can see for the largest threshold the success probability reaches

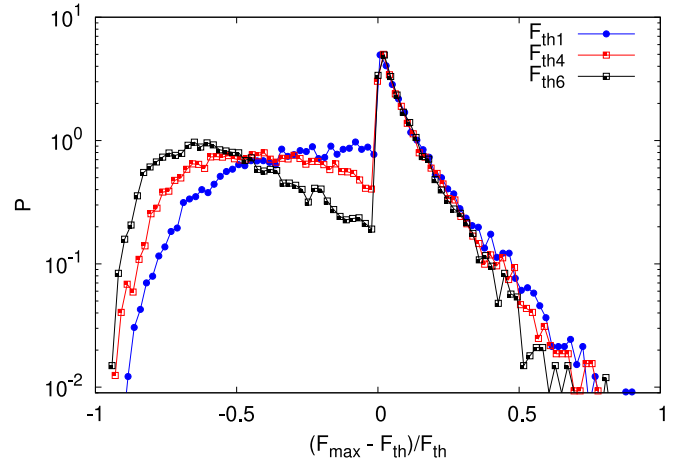


Fig. 8. PDF of the relative deviation between the maximal force and the threshold, conditioned to the different threshold values, F_{th1} , F_{th4} and F_{th6} . Data refer to the GNN-Cl tested on the database Cl-bal.

92% while for the smaller one is only 78,3% and similar discrepancies for the F_1 -score.

We remark that the number of instances with larger threshold are less than those with smaller threshold because of the condition of having at maximum one bond above threshold, and the larger the threshold the larger is the probability to have more than one bond

Table 3

Performances of the classifier model GNN-CI in the test-set CI-bal, with results disaggregated by the values of the threshold on the maximal tensile force. Col 1: threshold name; col 2: value of the threshold; col 3: total number of aggregate samples; col 4: number of broken aggregates; col 5 number of unbroken aggregates; col 6: total number of correct predictions; col 7: total number of wrong predictions; col 8: false negatives, or BCU, i.e. true Broken aggregates Classified as Unbroken; col 9: false positives or UCB, true Unbroken Classified as Broken; col 10: (unconditional) success probability; col 11: F_1 -score.

Th name	Th value	Tot	Broken	Unbroken	Correct	Wrong	BCU	UCB	%success	F_1 -score
1	1052	15067	7618	7449	11803	3264	2055	1209	78.3	0.80
2	1220	14906	7362	7544	12125	2781	1599	1182	81.3	0.82
3	1417	14945	7486	7459	12437	2508	1389	1119	83.2	0.84
4	1653	15245	7540	7705	13081	2164	1120	1044	85.8	0.86
5	1988	14931	7516	7415	13287	1644	827	817	88.9	0.89
6	2539	14906	7478	7428	13721	1185	572	613	92.0	0.92

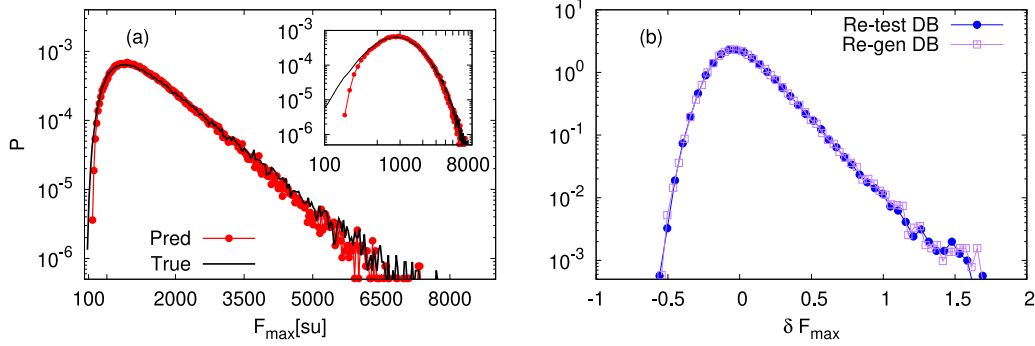


Fig. 9. (a) PDF of the maximal tensile force (black curve), and of the one predicted by the GNN-Re (red circles), data refer to the test-set Re-gen; (inset) the same but in log-log coordinates, to better appreciate differences in the PDF tails. (b) the PDF of the relative error between the predicted maximal tensile force and the ground-truth one, $\delta F_{\max} \equiv \frac{F_{\max}^{\text{pred}} - F_{\max}}{F_{\max}}$, using the GNN-Re in two different test-sets: data from Re-test (blue circles) and from Re-gen (pink squares) (see definitions in Section 3.2).

above it. Similar results are obtained in the test-set CI-rnd, with random thresholds. Here, however the presence of a larger number of wrong predictions for small thresholds is further amplified: the smaller the threshold the larger the probability to have $F_{\max} \approx F_{\text{th}}$ for unbroken aggregates. In other words, there are proportionally more samples which are difficult to handle for the GNN-CI (see Fig. 7).

Summarizing, we conclude that the smaller the threshold the larger the number of wrong predictions by the models, as confirmed by the F_1 -score.

6.2. Regression task

We now discuss the performances of the GNN-Re model on the test-sets Re-test and Re-gen (no threshold values, no selection of aggregates with single bond rupture). Fig. 9 (left) shows the predicted distribution of the maximal tensile force against the ground-truth one computed by the Stokesian dynamics in the test DB made of the aggregates not seen during the learning. The predicted PDF faithfully reproduces the true one, though some differences can be appreciated in the tails. In particular, we can see that in both the right and left tails (see inset) the PDF of the predicted $F_{\max}$ is slightly below the ground truth meaning that small/large $F_{\max}$ tend to be over/underestimated. This effect is further quantified in Fig. 9 (right), where we show distribution of the (signed) relative error of the predicted maximal tensile force from the true one, $\delta F_{\max} = \frac{F_{\max}^{\text{pred}} - F_{\max}}{F_{\max}}$ for both the test-sets Re-test and Re-gen. We observe that the distribution is not symmetric, i.e. for meaning that the model rather overestimates and underestimates the maximal tensile force. Moreover, we also observe that there are no appreciable differences between the results of the two test-sets, suggesting a good generalization capability of the regression model. To better disentangle the model behavior we look at the mean (signed) relative deviation, $\langle \delta F_{\max} | F_{\max} \rangle$, conditioned on the value of the maximal tensile force (see Fig. 10). We clearly observe that the GNN-Re tends to overestimate the force when this is small and underestimate it when this is large;

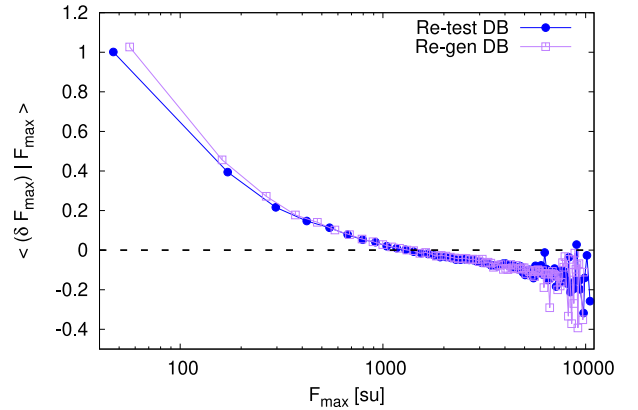


Fig. 10. Mean relative error between the predicted maximal tensile force and the true one, obtained from Stokesian Dynamics, $\langle \delta F_{\max} | F_{\max} \rangle$. The GNN-Re model is tested on two different test-sets: data from Re-test (blue circles) and from Re-gen (purple squares).

here, by *small* and *large*, we mean values of the maximal tensile force $F_{\max}$ which are smaller/larger than the peak of the distribution (these are the values with zero the mean deviation). We also note that the zero crossing of the signed deviations (lower curves) is not an indication of a small error, but of the fact that for this value of the force the model can overestimate and underestimate the force with equal probability. Furthermore, one can see that on average the modulus of the relative error is larger for smaller $F_{\max}$.

Finally, in Fig. 11, we compare the prediction by the GNN-Re model of the maximal tensile force distribution in the datasets built up with aggregates having the breakup of a single bond for a given threshold value, in the test-sets with fixed thresholds (CI-imb) or random (CI-rnd). As one can see the effect of the threshold on the PDF, signaled by the

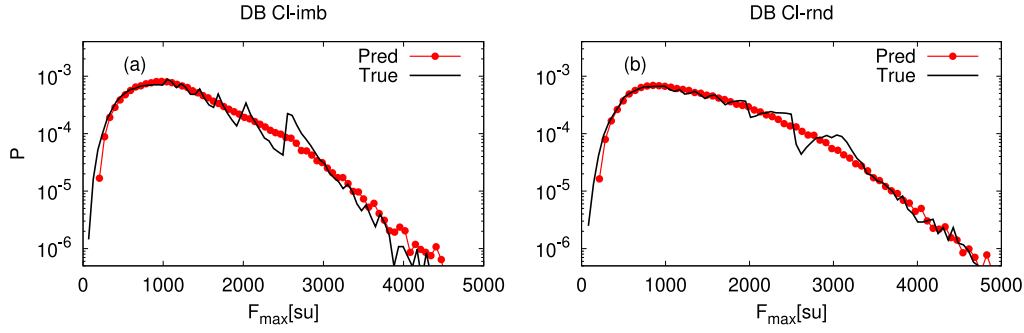


Fig. 11. (a) Comparison of the probability density functions of the ground-truth maximal force (continuous curve), and the one predicted by the GNN-Re in the test-set CI-imb (purple squares); (b) the same but testing the GNN-Re on data of the DB CI-rnd, with random thresholds (filled circles). The ‘spikes’ in the PDF of the ground-truth PDF correspond to the threshold values.

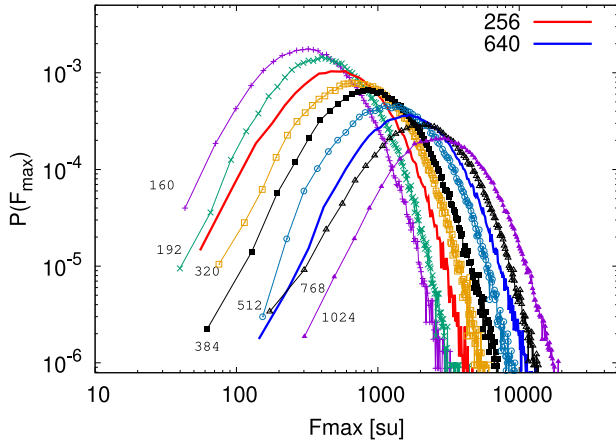


Fig. 12. PDF of the maximal tensile force $F_{\max}$ for aggregates of different sizes (or families). Each family is defined by the number of primary particles composing the aggregates. Lines with symbols refer to families of the new training database CI_{NM} (aggregates made with $N_M = 160, 192, 320, 512, 768, 1024$), while continuous lines refer to families of the new heterogeneous testing database, CI_{heter} (aggregates made with $N_M = 256, 640$).

jumps in the PDFs, is smoothed out by the regressor model. Eventually, this can be cured by training the regressor directly on the database used by the GNN-CI and providing it the information on the threshold.

6.3. Model heterogeneous generalization to different sizes

Evaluating the ability of data-driven models to generalize to unseen data is essential, as it indicates that they capture the underlying relationships rather than simply memorizing the training data. In the previous section, we tested the models’s capability to generalize to unseen geometrical and velocity gradient realizations. Here we want to push this further and consider the model ability to generalize to aggregates of different sizes, namely aggregates with variable number of primary particles. As it is clear from Fig. 12, aggregates of different sizes are characterized by similar distributions of the maximal tensile forces, but over quite different ranges of values. This has an important consequence: a GNN must be trained on a dataset containing aggregates of heterogeneous sizes to learn the size-dependent tensile force distributions; once trained, it can generalize to aggregate sizes not encountered during training. This is a very robust test of GNN classifier ability to perform well in more complex suspensions.

For this goal, we built new databases, for training and testing respectively. For the GNN classifier model training, we consider a well balanced dataset composed of aggregates with variable number

of primary particles $N_M = 160, 192, 320, 384, 512, 768, 1024$ with 40 different geometries for each N_M .

We collect 60 000 samples for each N_M for a total of 420 000 samples, half broken and half unbroken. This is called CI_{NM} , and it is further divided in two subsets, with 80% of samples used for training the GNN classifier and 20% for standard testing on aggregate families (or sizes) statistically homogeneous with those seen during the training.

To test the GNN-classifier model on aggregates of families never seen during the training, we consider a dataset containing aggregates with different number of primary particles, namely $N_M = 256, 640$; this database, called CI_{heter} and containing 120 000 realizations, is also well balanced among the two aggregate families, and in broken/unbroken samples.

Given the strong dependence of the maximal tensile force on the aggregate size, the training of the new GNN Classifier is done selecting random thresholds for each aggregate family; N_M -specific thresholds are extracted from the cumulative PDF of the maximal tensile force, and vary in the range [0.3–0.9] of the CDF.

The results are summarized in Table 4: first we observe that the GNN-Classifer when tested on aggregate families seen during the training (but clearly with different velocity gradients and rotation realizations) behaves very well with an average success probability of 87%; second, we observe that the GNN-Classifer maintains such performances also when operating on aggregates of sizes and maximal tensile force distributions never seen during the training. Moreover, the success probability and F_1 -score are basically independent of N_M , confirming the robustness of the performances. These results point to the strong capability of learning and generalizing of our model.

7. Conclusion

We have investigated the ability of Graph Neural Networks to predict the rupture of isostatic aggregates by hydrodynamical stresses in three-dimensional turbulent suspensions. While we do not *interpret* our models (Mehlig, 2021) - something that goes well beyond the limits of our work-, we can assess their validity and investigate their failures.

In particular, we have proposed two distinct implementations of the GNNs, targeting classification or regression tasks. We showed that we generally obtain very good accuracy in the breakup prediction, and that models tend to fail for force values very close to the rupture thresholds. Clearly, the threshold region is the one where making correct predictions is more difficult. In applications where the goal is to avoid rupture, this may not be a problem since the GNN prediction would represent an upper bound to a conservative estimation; for applications targeting the formation of small from large aggregates, GNN results should be interpreted with additional care.

Concerning generalization of model behavior, we showed that the implemented GNNs exhibit very good accuracy when the test-set characteristics differ from those seen during the training. This applies when

Table 4

Col 1: name of the test database; col 2: aggregate family type, defined in terms of the number of primary particles N_M within the aggregate; col 3: total number of aggregate samples; col 4: number of broken aggregates; col 5: number of unbroken aggregates; col 6: total number of correct predictions; col 7: total number of wrong predictions; col 8: false negatives, or BCU; col 9: false positives or UCB; col 10: (unconditional) success probability, giving the correct prediction over the total number of samples; col 11: F₁-score, giving the harmonic mean of precision and recall. Note the different composition of the two databases: Cl_{NM} is a test dataset containing aggregate families seen during training; Cl_{heter} is a test dataset containing aggregate families never seen during training, and it is thus used to fully test generalization ability of the GNN-Classifier.

DB	N_M	Tot	Broken	Unbroken	Correct	Wrong	BCU	UCB	%success	F ₁ -score
Cl_{NM}	160	12 064	6028	6036	10 453	1611	479	1132	86.6	0.87
Cl_{NM}	192	12 043	5976	6067	10 471	1572	400	1172	86.9	0.87
Cl_{NM}	320	11 835	5939	5896	10 219	1616	439	1177	86.3	0.87
Cl_{NM}	384	11 937	5944	5933	10 396	1541	470	1071	87.0	0.87
Cl_{NM}	512	12 020	6018	6002	10 646	1374	334	1040	88.5	0.89
Cl_{NM}	768	12 019	5950	6069	10 589	1430	418	1012	88.1	0.88
Cl_{NM}	1024	12 082	5988	6094	10 681	1401	423	978	88.4	0.88
Cl_{NM}	all	84 000	41 843	42 157	73 455	10 545	2963	7582	87.4	0.88
Cl_{heter}	256	60 000	30 000	30 000	52 480	7520	1706	5814	87.4	0.88
Cl_{heter}	640	60 000	30 000	300 000	52 523	7477	2129	5384	87.5	0.88

considering random threshold values for the rupture for the classifier model, and when considering new geometrical shapes for the regression model. Similarly, very good performances have been obtained when training the model on heterogeneous in sizes aggregates, and testing it on aggregates of sizes never seen during the training.

Our study opens the way to further applications in the field of colloidal suspensions, or microplastics modeling in complex flows. A promising avenue for future studies is training a GNN model with full temporal histories of stress and forces. By doing so, we could produce synthetic trajectories of aggregate dynamics (see e.g. Li et al. (2024)), complemented by the tensile forces, and give dynamical predictions, substituting numerical integration in flows at different/higher Reynolds numbers.

Another interesting direction is to develop more sophisticated GNN architectures, exploiting for instance attention mechanisms not only to predict the maximal tensile force value, F_{max} , but also the bond where this is realized, or to classify the rupture and identify the broken bond. This would eventually speed up the computation of breakup events and the aggregate size distribution in numerical simulations with multiple aggregates. A long-term goal targets the modeling of the interplay of aggregation/fragmentation processes by data-driven methods.

CRedit authorship contribution statement

M. Buzzicotti: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **M. Cencini:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **G. Cimini:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **M. Vanni:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **A.S. Lanotte:** Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

M. C. and A. L. acknowledge support by the European Union Next Generation plan, through the Italian Ministry of University and Research (MUR) Piano Nazionale di Ripresa e Resilienza (PNRR), project title: 'CN00000013 – National Centre for HPC, Big Data and Quantum Computing'. M.B. acknowledge support by the Fondo Italiano per la Scienza 2022–2023 (FIS2) CUP E53C24003760001 A.L. is grateful to Gautier Verhille and Michael Wilczek for useful discussions.

Appendix A

The datasets listed in Table 2 have been prepared as follows.

Turbulent velocity gradients: We perform Direct Numerical Simulations of Homogeneous and Isotropic, forced three-dimensional turbulence for the fluid velocity $\mathbf{v}(\mathbf{x}, t)$. Once the Eulerian flow is statistically stationary, we integrate the Lagrangian trajectories of tracer particles, $d\mathbf{X}/dt = \mathbf{v}(\mathbf{X}(t), t)$. Along the tracer trajectories, we store all components of the fluid velocity gradient tensor at the tracer position, $\partial_i v_j(\mathbf{X}(t))$. Then along each trajectory we select 2048 velocity gradient realizations, sufficiently separated in time (about one large-scale eddy-turn-over-time) such that they are uncorrelated. So doing, we have built up a collection of independent and identically distributed (with the correct turbulent statistics) velocity gradient realizations. We remark that DNS are needed to obtain the intermittent velocity gradient statistics that is observed in three-dimensional turbulence (Frisch, 1995).

Aggregates: We consider aggregates with the same fractal dimension $D_F = 1.9$ and number of primary particles $N_M = 384$, but different geometrical arrangements. We take each aggregate geometry and rotate it by a random angle in the 3D space, and store independent orientations of the same aggregate. At the end, we have a collection of aggregates with independent orientations and independent realizations of the turbulent velocity gradients. In this set-up, any time is equivalent. For each of these, we compute the flow induced forces/torques within the Stokesian Dynamics approach. These data are stored and constitute the starting DB.

Breakup: Having computed the tensile force distribution over all samples with the Stokesian Dynamics, we identify rupture threshold values. Then to each sample aggregate it is associated a rupture threshold: since we know from SD the maximal tensile force for the specific aggregate sample and we have a rupture value threshold, we know if the aggregate wrt that threshold is a broken one (provided it has one, and only one, internal bond with tensile force above the threshold) or unbroken one (provided it has all internal bonds with the tensile forces below threshold).

Data availability

Data will be made available on request.

References

- Babler, M.U., Biferale, L., Brandt, L., Feudel, U., Guseva, K., Lanotte, A., Marchioli, C., Picano, F., Sardina, G., Soldati, A., Toschi, F., 2015. Numerical simulations of aggregate breakup in bounded and unbounded turbulent flows. *J. Fluid Mech.* 766, 104–128.
- Balachandar, S., Eaton, J., 2010. Turbulent dispersed multiphase flow. *Annu. Rev. Fluid Mech.* 42, 111–133.
- Barwey, S., Pal, P., Patel, S., Balin, R., Lusch, B., Vishwanath, V., Maulik, R., Balakrishnan, R., 2025. Mesh-based super-resolution of fluid flows with multiscale graph neural networks. *Comput. Methods Appl. Mech. Engrg.* 443, 118072.
- Becker, V., Schlauch, E., Behr, M., Briesen, H., 2009. Restructuring of colloidal aggregates in shear flows and limitations of the free-draining approximation. *J. Colloid Interface Sci.* 339 (2), 362–372.
- Benzi, R., Biferale, L., Calzavarini, E., Lohse, D., Toschi, F., 2009. Velocity-gradient statistics along particle trajectories in turbulent flows: The refined similarity hypothesis in the Lagrangian frame. *Phys. Rev. E* 80 (6), 066318.
- Bilodeau, R., Bousfield, D., 1998. Shear-thinning predictions from particle motion modeling. *J. Rheol.* 42, 743–764.
- Brady, J.F., Bossis, G., 1988. Stokesian dynamics. *Annu. Rev. Fluid Mech.* 20, 111–157.
- Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P., 2017. Geometric deep learning: Going beyond euclidean data. *IEEE Signal Process. Mag.* 34 (4), 18–42. <http://dx.doi.org/10.1109/MSP.2017.2693418>.
- Chen, S., Li, S., 2020. Collision-induced breakage of agglomerates in homogenous isotropic turbulence laden with adhesive particles. *J. Fluid Mech.* 902, A28. <http://dx.doi.org/10.1017/jfm.2020.582>.
- Corso, G., Stark, H., Jegelka, S., Jaakkola, T., Barzilay, R., 2024. Graph neural networks. *Nat. Rev. Meth. Prim.* 4 (1), 17. <http://dx.doi.org/10.1038/s43586-024-00294-7>.
- De Bona, J., Lanotte, A., Vanni, M., 2014. Internal stresses and breakup of rigid isotactic aggregates in homogeneous and isotropic turbulence. *J. Fluid Mech.* 755, 365.
- Derksen, J., 2013. Direct simulations of aggregates in homogeneous isotropic turbulence. *Acta Mech.* 10, 2415–2424.
- Eggersdorfer, M., Kadau, D., Herrmann, H., Pratsinis, S., 2010. Fragmentation and restructuring of soft-agglomerates under shear. *J. Colloid Interface Sci.* 342, 261–268.
- Faneli, M., Feke, D., Manas-Zloczower, I., 2006. Prediction of the dispersion of particle clusters in the nano-scale - Part I: Steady shearing responses. *Chem. Eng. Sci.* 61, 473–488.
- Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognit. Lett.* 27 (8), 861–874.
- Filiatraut, A.N., Mianroodi, J.R., Siboni, N.H., Zanjani, M.B., 2023. Predicting micro/nanoscale colloidal interactions through local neighborhood graph neural networks. *J. Appl. Phys.* 134 (23), 234702. <http://dx.doi.org/10.1063/5.0175062>.
- Frisch, U., 1995. *Turbulence: The Legacy of An Kolmogorov*. Cambridge University Press.
- Frungieri, G., Boccardo, G., Buffo, A., Karimi-Varzaneh, H., Vanni, M., 2022. CFDDEM characterization and population balance modelling of a dispersive mixing process. *Chem. Eng. Sci.* 260, 117859.
- Frungieri, G., Vanni, M., 2021. Aggregation and breakup of colloidal particle aggregates in shear flow: A combined Monte Carlo - Stokesian dynamics approach. *Powder Tech.* 388, 357–370.
- Gao, R., Jaiman, R.K., 2024. Predicting fluid-structure interaction with graph neural networks. *Phys. Fluids* 36, 013622.
- Gavi, E., Marchisio, D., Barresi, A., Olsen, M., Fox, R., 2010. Turbulent precipitation in micromixers: CFD simulation and flow field validation. *Chem. Eng. Res. Des.* 88, 1182–1193.
- Hamilton, W., Ying, Z., Leskovec, J., 2017. Inductive representation learning on large graphs. *Adv. Neural Inf. Process. Syst.* 30.
- Harada, S., Tanaka, R., Nogami, H., Sawada, M., 2006. Dependence of fragmentation behavior of colloidal aggregates on their fractal structure. *J. Colloid Interface Sci.* 301 (1), 123–129.
- Harshe, Y., Lattuada, M., 2012. Breakage rate of colloidal aggregates in shear flow through Stokesian dynamics. *Langmuir* 28, 283–292.
- Harshe, Y., Lattuada, M., Soos, M., 2011. Experimental and modeling study of breakage and restructuring of open and dense colloidal aggregates. *Langmuir* 27, 5739–5752.
- Higashitani, K., Iimura, K., Sanda, H., 2001. Simulation of deformation and breakup of large aggregates in flows of viscous fluids. *Chem. Eng. Sci.* 56 (9), 2927–2938.
- Jarvis, P., Jefferson, B., Gregory, J., Parsons, S.A., 2005. A review of floc strength and breakage. *Water Res.* 39 (14), 3121–3137.
- Johnson, K.L., 1985. *Contact Mechanics*. Cambridge University Press.
- Kaveh, K., Malcherek, A., 2025. Theoretical framework for modeling flocculation in cohesive sediments with variable yield strength. *Water Res.* 282, 123593.
- Khalifa, A., Breuer, M., 2023. Data-driven ANN approach for binary agglomerate collisions including breakage and agglomeration. *Chem. Engin. Res. Des.* 195, 14–27.
- Khalifa, A., Breuer, M., 2024. Evaluation of an efficient data-driven ANN model to predict agglomerate collisions within Euler-Lagrange simulations. *Comput. & Fluids* 269, 106119.
- Kusters, K.A., Wijers, J.G., Thoenes, D., 1997. Aggregation kinetics of small particles in agitated vessels. *Chem. Eng. Sci.* 52 (1), 107–121.
- Lattuada, M., Wu, H., Morbidelli, M., 2003. A simple model for the structure of fractal aggregates. *J. Colloid Interface Sci.* 268 (1), 106–120.
- Leonelli, A.D., Widmer, L., Meiburg, E., 2026. Aggregation, breakup, and size-dependent transport in a turbulent channel flow with cohesive particles. <https://arxiv.org/abs/2604.20028>.
- Li, T., Biferale, L., Bonaccorso, F., Scarpolini, M.A., Buzzicotti, M., 2024. Synthetic Lagrangian turbulence by generative diffusion models. *Nat. Mach. Intell.* 6, 393–403.
- Ma, Z., Ye, Z., Pan, W., 2022. Fast simulation of particulate suspensions enabled by graph neural network. *Comput. Methods Appl. Mech. Engrg.* 400, 115496. <http://dx.doi.org/10.1016/j.cma.2022.115496>.
- Manas-Zloczower, I., 2012. *Mixing and compounding of polymers: Theory and practice*. Carl Hanser Verlag, Munich.
- Mehlig, B., 2021. *Machine Learning with Neural Networks: An Introduction for Scientists and Engineers*. Cambridge University Press.
- Meneveau, C., Sreenivasan, K.R., 1991. The multifractal nature of turbulent energy dissipation. *J. Fluid Mech.* 224, 429–484.
- Qi, Y., Shin, S., Coletti, F., 2025. Breakup and coalescence of particle aggregates at the interface of turbulent liquids. *J. Fluid Mech.* 1021, A49.
- Reiser, P., Neubert, M., Eberhard, A., Torresi, L., Zhou, C., Shao, C., Metni, H., van Hoesel, C., Schopmans, H., Sommer, T., Friederich, P., 2022. Graph neural networks for materials science and chemistry. *Commu. Mater.* 3, 93.
- Ren, Z., Harshe, Y., Lattuada, M., 2015. Influence of the potential well on the breakage rate of colloidal aggregates in simple shear and uniaxial extensional flows. *Langmuir* 31, 5712–5721.
- Ruan, X., Chen, S., Li, S., 2020. Structural evolution and breakage of dense agglomerates in shear flow and Taylor-green vortex. *Chem. Eng. Sci.* 211, 115261.
- Saha, D., Bäbler, M., Holzner, M., Soos, M., Lüthi, B., Liberzon, A., Kinzelbach, W., 2016. Breakup of finite-size colloidal aggregates in turbulent flow investigated by three-dimensional (3D) particle tracking velocimetry. *Langmuir* 32, 55–65.
- Saha, D., Soos, M., Lüthi, B., Holzner, M., Liberzon, A., Bäbler, M., Kinzelbach, W., 2014. Experimental characterization of breakage rate of colloidal aggregates in axisymmetric extensional flow. *Langmuir* 30, 14385–14395.
- Sanchez-Gonzalez, A., Godwin, J., Pfaff, T., Ying, R., Leskovec, J., Battaglia, P., 2020. Learning to simulate complex physics with graph networks. In: III, H.D., Singh, A. (Eds.), *Proceedings of the 37th International Conference on Machine Learning*. In: *Proceedings of Machine Learning Research*, vol. 119, PMLR, pp. 8459–8468.
- Sawford, B.L., 1991. Reynolds number effects in Lagrangian stochastic models of turbulent dispersion. *Phys. Fluids A* 3 (6), 1577–1586.
- Saxena, A., Kroll-Rabotin, J.S., Sanders, R., 2022. Numerical investigation of the respective roles of cohesive and hydrodynamic forces in aggregate restructuring under shear flow. *J. Colloid Interface Sci.* 608, 355–365.
- Saxena, A., Kroll-Rabotin, J.S., Sanders, R., 2023. Role of flow inertia in aggregate restructuring and breakage at finite Reynolds numbers. *Langmuir* 39, 10066–10078.
- Saxena, A., Kroll-Rabotin, J.S., Sanders, R., 2025. Exposure of fractal aggregates to accelerating flows at finite Reynolds numbers. *Int. J. Multiph. Flow* 182, 105018.
- Scarselli, F., Gori, M., Tsoi, A., Hagenbuchner, M., Monfardini, G., 2009. The graph neural network model. *IEEE Trans. Neural. Netw.* 20 (1), 61–80.
- Schutte, K., Portela, L., Twerda, A., Henkes, R., 2018. Formation and break-up of rigid agglomerates in turbulent channel and pipe flows. *J. Fluid Mech.* 857, 539–561.
- Sergeev, P., Kiselev, D., Makarov, I., 2024. Graph neural networks for complex fluid simulations. In: *2024 IEEE 28th International Conference on Intelligent Engineering Systems. INES*, pp. 000145–000152. <http://dx.doi.org/10.1109/INES63318.2024.10629094>.
- Serhani, A., Xing, V., Dupuy, D., Lapeyre, C., Staffelbach, G., 2024. Graph and convolutional neural network coupling with a high-performance large-eddy simulation solver. *Comput. & Fluids* 278, 106306.
- Seto, R., Botet, R., Briesen, H., 2011. Hydrodynamic stress on small colloidal aggregates in shear flow using Stokesian dynamics. *Phys. Rev. E* 84, 041405.
- Thomas, A., Durmaz, A.R., Alam, M., Gumbsch, P., Sack, H., Eberl, C., 2023. Materials fatigue prediction using graph neural networks on microstructure representations. *Sci. Rep.* 13, 12562.
- Vanni, M., 2015. Accurate modelling of flow induced stresses in rigid colloidal aggregates. *Comp. Phys. Commu.* 192, 70–90.
- Vanni, M., Gastaldi, A., 2011. Hydrodynamic forces and critical stresses in low-density aggregates under shear flow. *Langmuir* 27, 12822–12833.

- Vasquez Giuliano, L., Buffo, A., Vanni, M., Lanotte, A., Arima, V., Bianco, M., Baldassarre, F., Frungieri, G., 2022. Response of shear-activated nanotherapeutic particles in a clot-obstructed blood vessel by CFD-DEM simulations. *Can. J. Chem. Eng.* 100, 3562–3572.
- Vu, D.T.S., Nguyen, T.M., Ren, W., 2025. Learning and predicting dynamics of compositional multiphase mixtures using Graph Neural Networks. *J. Comput. Phys.* 529, 113851.
- Yao, Y., Capecelatro, J., 2021. Deagglomeration of cohesive particles by turbulence. *J. Fluid Mech.* 911, A10.
- Yu, M., Yu, X., Mehta, A., Manning, A., Khan, F., Balachandar, S., 2023. Persistent reshaping of cohesive sediment towards stable flocs by turbulence. *Sci. Rep.* 13, 1760.
- Zaccone, A., Soos, M., Lattuada, M., Wu, H., Bäbler, M.U., Morbidelli, M., 2009. Breakup of dense colloidal aggregates under hydrodynamic stresses. *Phys. Rev. E* 79, 061401.
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M., 2020. Graph neural networks: A review of methods and applications. *AI Open* 1, 57–81. <http://dx.doi.org/10.1016/j.aiopen.2021.01.001>.