1. ABOUT THE DATASET

--------------------

Title: Data file outputs per simulation run

Creator(s): Kalila Rohana Cook

Organisation(s): University of Leeds

Rights-holder(s): Unless otherwise stated, Copyright 2022 University of Leeds

Publication Year: 2023

Description: This dataset consists of raw data file outputs from each simulation run. The *‘VolumeFractions'* folder contains subfolders with this repeated data set type but with parameters varying for each volume fraction/reaction probability/simulation run number.

Each data set is found in directory path: *‘VolumeFractions’ > ‘{x}percentVF’ > ‘R{y}’ > ‘Run{z}’*, where *{x}percentVF* = x% monomer volume fraction, *R{y}* = monomer-monomer reaction probability of y% and *Run{z}* = simulation repeat number, with z ranging from 1-10.

Each data set contains files for

* box-counting data (‘BoxCountData.csv’)
	+ box-covering data output with columns headed 'r' and 'num' (corresponding to 'r', measuring cubes of unit size, and 'O(r)', the occupation number, in the paper). Each file contains up to 4 sets of appended box-covering data, each headed in this way. The 3rd to last set is for the percolation cluster at the percolation point, the 2nd to last is for the full system at the percolation point and the last set is for the whole system at the end, fully crosslinked state. Exemplar fitted data is presented in Fig. S4. Box-covering data was fitted and processed to extract the fractal dimensions in Fig. 5.
* cluster sizes (‘ClusterSizes.csv’)
	+ measures of cluster size over time. Columns are headed ‘T’ (time step), ‘Mean’ (average cluster size), ‘Largest’ (largest cluster size) and ‘PercolSize’ (size of the percolation cluster). T vs PercolSize data (averaged over 10 runs for each parameter set) is presented in Fig. 2b and 2c.
* gel point time and position coordinates (‘GelPoint\_T{GEL\_TIME}.csv’)
	+ monomer position coordinates in 3D at the percolation point. The time, {GEL\_TIME}, in the filename corresponds to the exact timestep that percolation is first detected (percolation time, τ, in the paper). These values were used to produce Fig. 3b and Fig. 4a in the paper and Fig. S1 in ESI. The coordinates were also used to produce the percolation snapshots of Fig. 1a(2), Fig. 3a, Fig. 4a and Fig. 5a and b.
* position coordinates for initial monomer positions (‘InitialParticlePositions.csv’)
	+ monomer position coordinates in 3D at the start of the simulation, T=0. These coordinates were used to produce the initial state snapshot of Fig. 1a(1).
* position coordinates for updated monomer positions (‘UpdatedParticlePositions.csv’)
	+ monomer position coordinates in 3D at the end of the simulation at the fully crosslinked state. These coordinates were used to produce the final state snapshots of Fig. 1a(3), Fig. 2a, and Fig. 5c. They were also used to produce the pore distribution snapshots of Fig. 6a and c.
* cluster mass distribution (‘ClusterMassDistributions.csv’)
	+ cluster size distributions at the percolation point. Columns are headed ‘Index’ for the cluster index number stored in memory and ‘Mass’ for the size of the cluster in terms of number of monomers. A cumulative frequency plot of this data for 10 runs each of a 6% monomer volume fraction system at reaction probabilities of 100 and 0.2% data is presented in Fig. S2.
* supercomputer job script output (‘{JOB\_NAME}.sh.o{JOBID}’)
	+ Supercomputer job script output. {JOB\_NAME} corresponds to varying volume fraction/reaction probability filename. '{JOBID}' corresponds to varying assigned supercomputer job script ID.

Cite as: Cook, Kalila (2023): ‘Data file outputs per simulation run’. University of Leeds. Dataset 1. <https://doi.org/10.5518/1323>.

Related publication: Cook, K. R., Head, D., & Dougan, L. (2023). Modelling Network Formation in Folded Protein Hydrogels by Cluster Aggregation Kinetics. Soft Matter (Accepted). https://doi.org/10.1039/D3SM00111C.