Graph theory: From simulation to information

Clément Lagisquet, Julian C. Shillcock, Jérémy Alexandre, Laurent Vuillon, View John H. Ipsen

LAMA, Université Savoie Mont Blanc

3 mai 2022

Overview

From biology to graph theory

Clustering coefficient

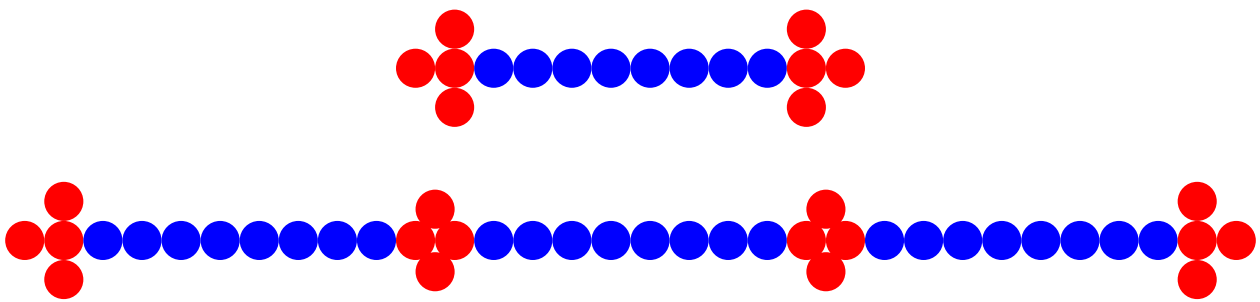
Fluidity

Overview

This talk is based on : *Model biomolecular condensates have heterogeneous structure quantitatively dependent on the interaction profile of their constituent macromolecules*, Julian C. Shillcock, CL, Jérémy Alexandre, Laurent Vuillon, View John H. Ipsen

Polymers

In order to understand how polymers behave, we simulate them using molecular dynamics.
Polymer can have two or more binding sites (here in red), which are the only area where they have interactions and the blue parts are flexible.

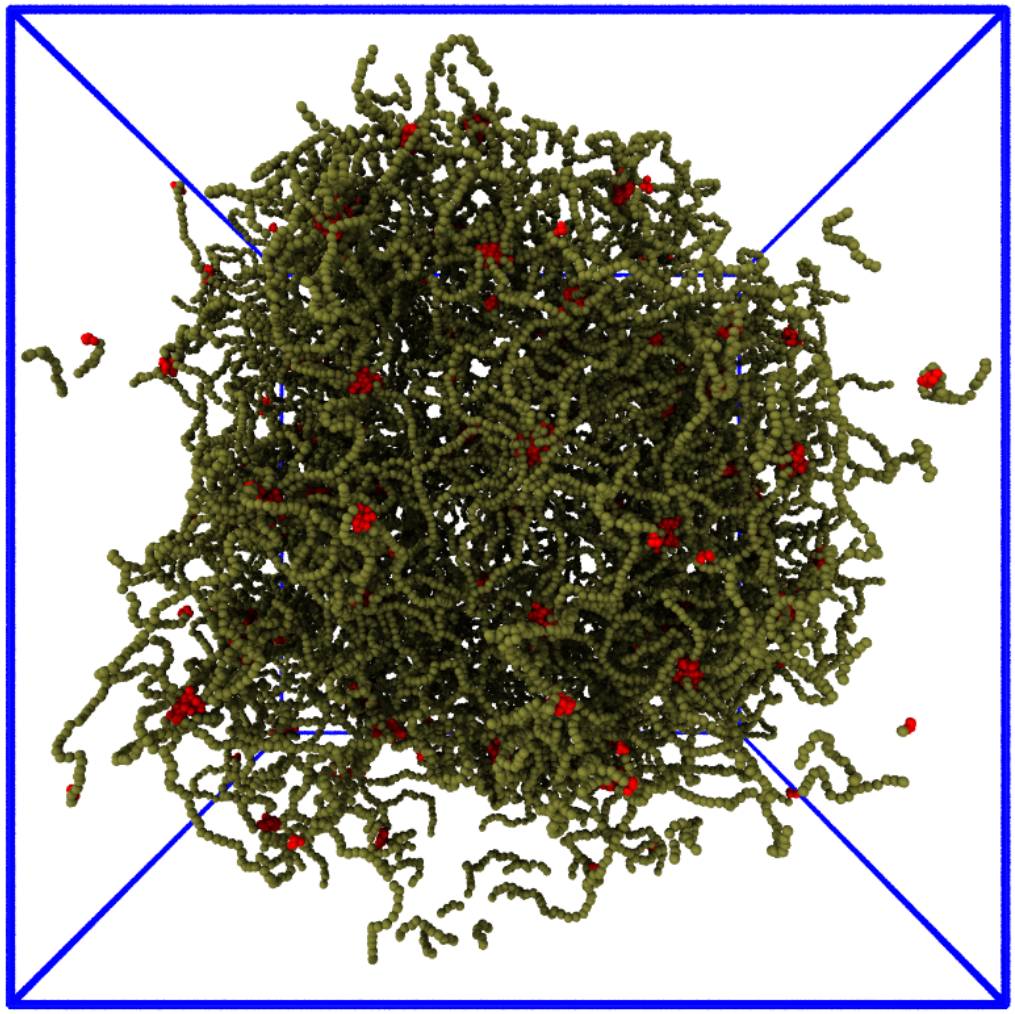


Polymers

We have three parameters :

- the concentration, i.e. the number of polymers ;
- the lengths between two consecutive binding sites ;
- an affinity parameter, which says how much binding sites will be attracted to each other.

Droplets



What to say about that ?

It ressembles a plate of bolognese pasta ?

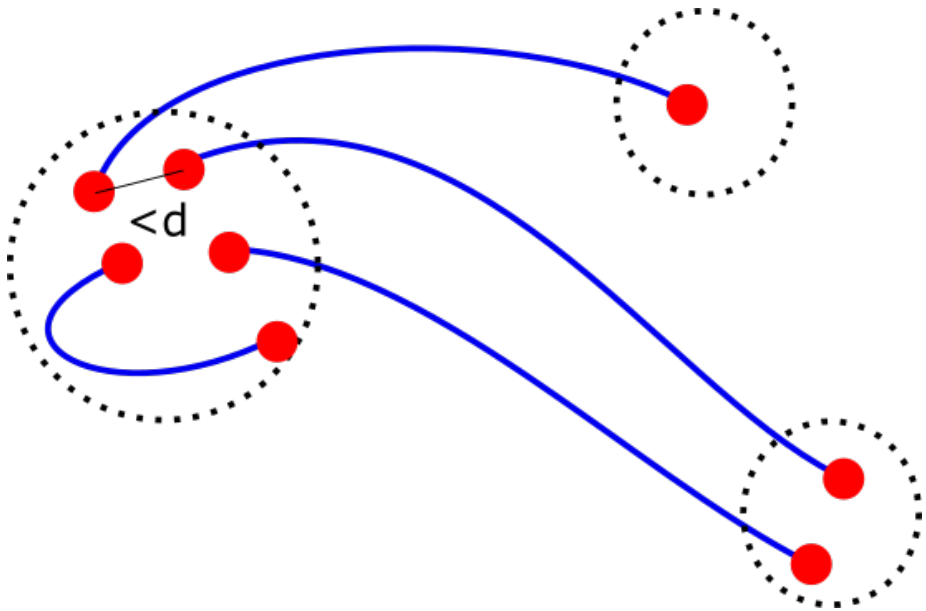
What to say about that ?

How can we extract some meanings from that kind of objects ?
The two questions that we will address in this talk are :

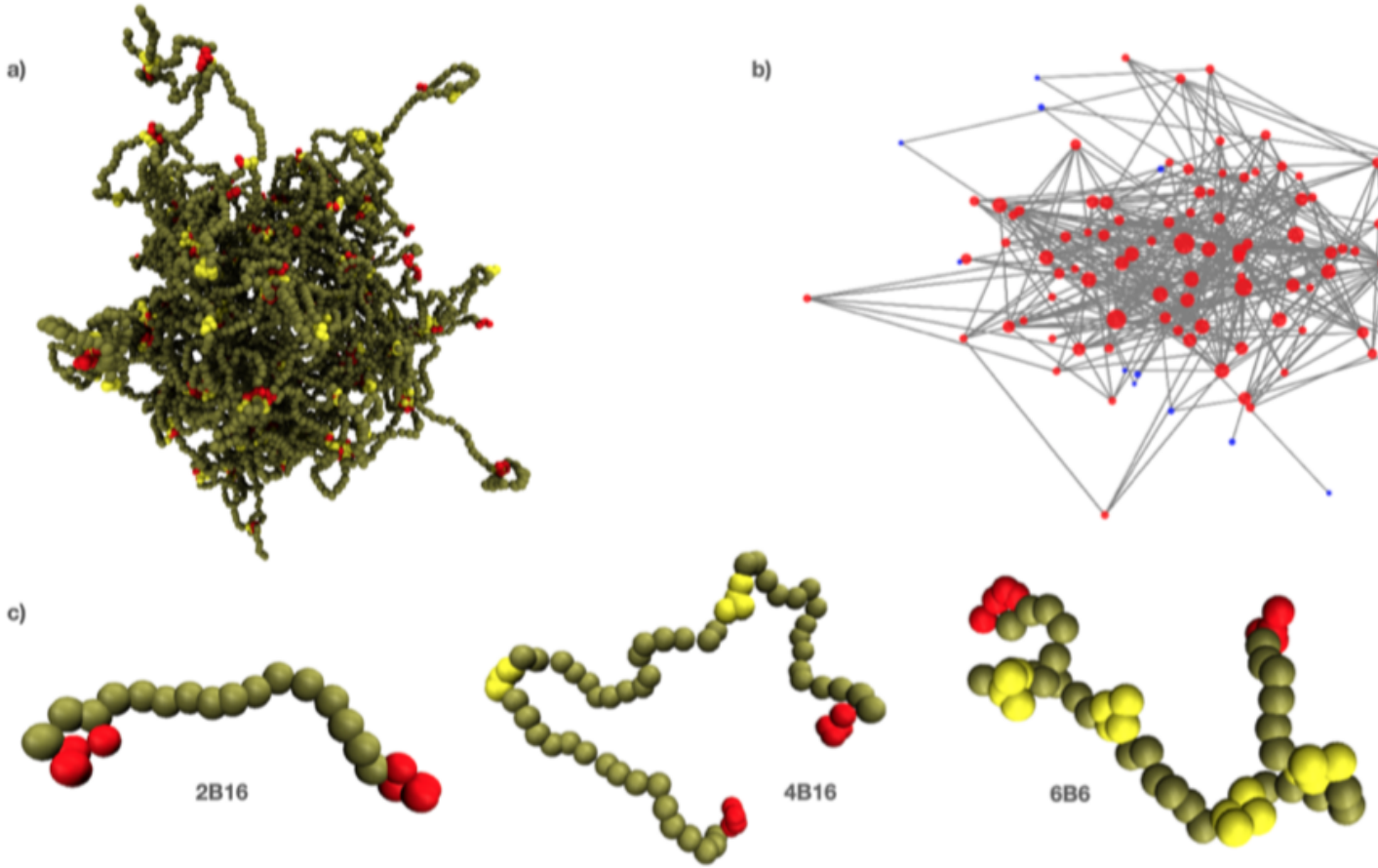
- how crowded are the polymers around junctions ?
- how much do they diffuse in the droplet ?

Here comes the graphs

There is a "natural" transition to a graph.
Where you clusters the binding sites if they are at a certain distance of each other that create our vertices and the edges are the connections made by the polymers between those clusters. And the weight of these edges are given by the number of links spanning the two clustered vertices.



Here comes the graphs



Clustering coefficient

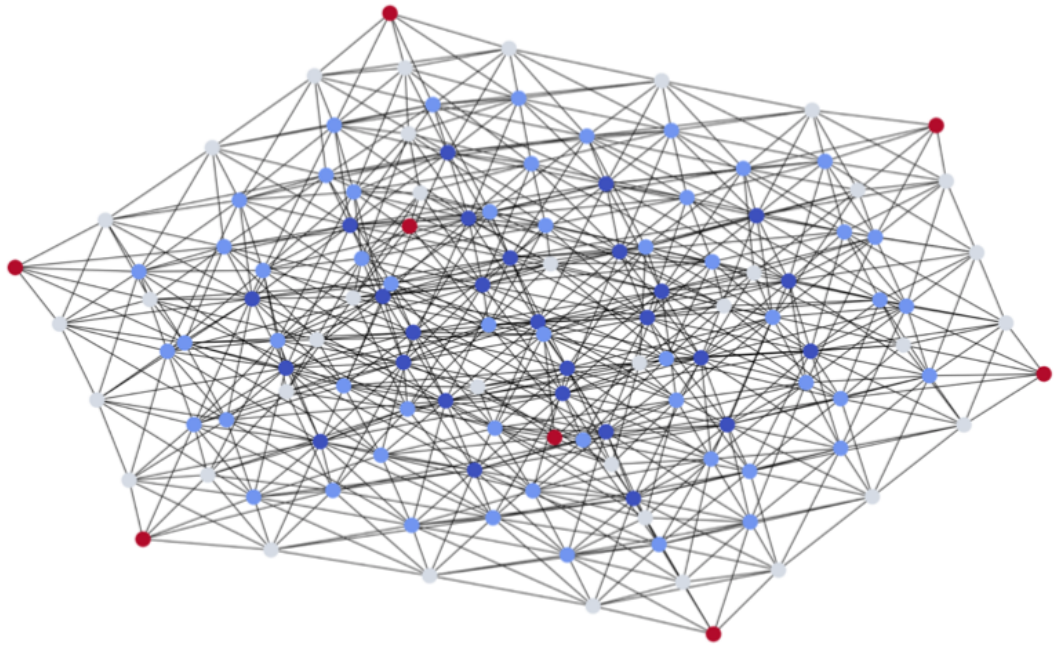
Let $G(V, E)$ be a graph and A its adjacency matrix then the clustering coefficient of a node n_i is define as

$$c_i = \frac{1}{k_i(k_i - 1)} \sum_{j,k} A_{i,j} A_{j,k} A_{k,i},$$

where k_i is the degree of n_i.

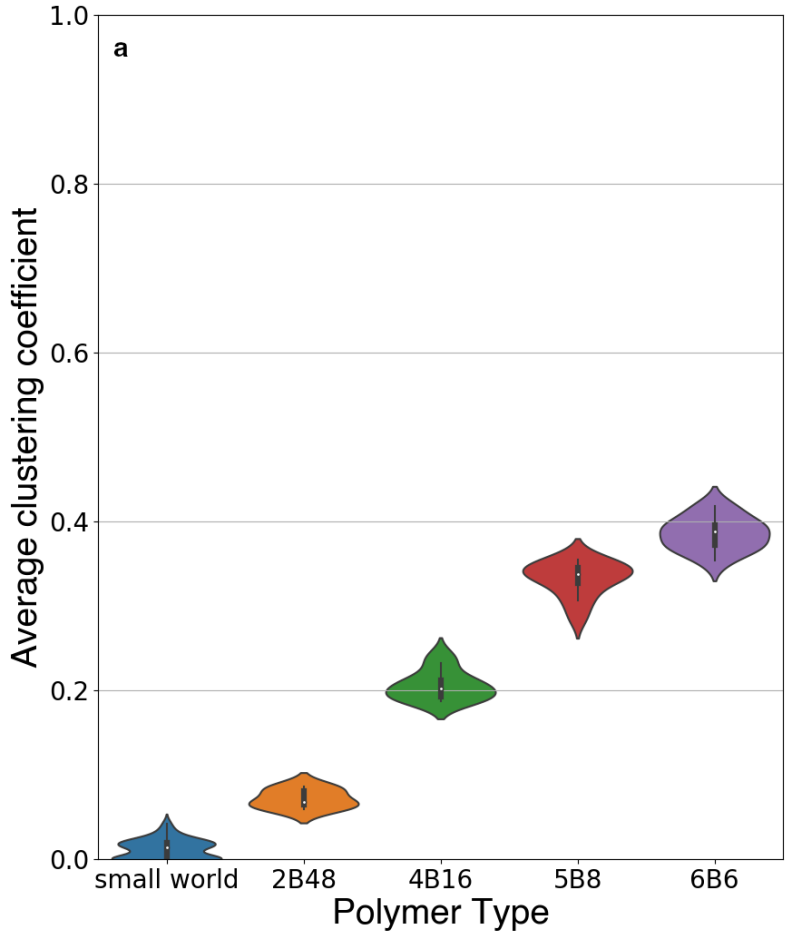
The clustering coefficient of G is $\frac{1}{|V|} \sum_i c_i$.

Example



- vertice : $4/5$
- edge : $7/12 \approx 0.58$
- facet : $6/13 \approx 0.46$
- core : $20/51 \approx 0.39$

Results


a
1.0
0.8
0.6
0.4
0.2
0.0
Average clustering coefficient
small world
2B48
4B16
5B8
6B6
Polymer Type

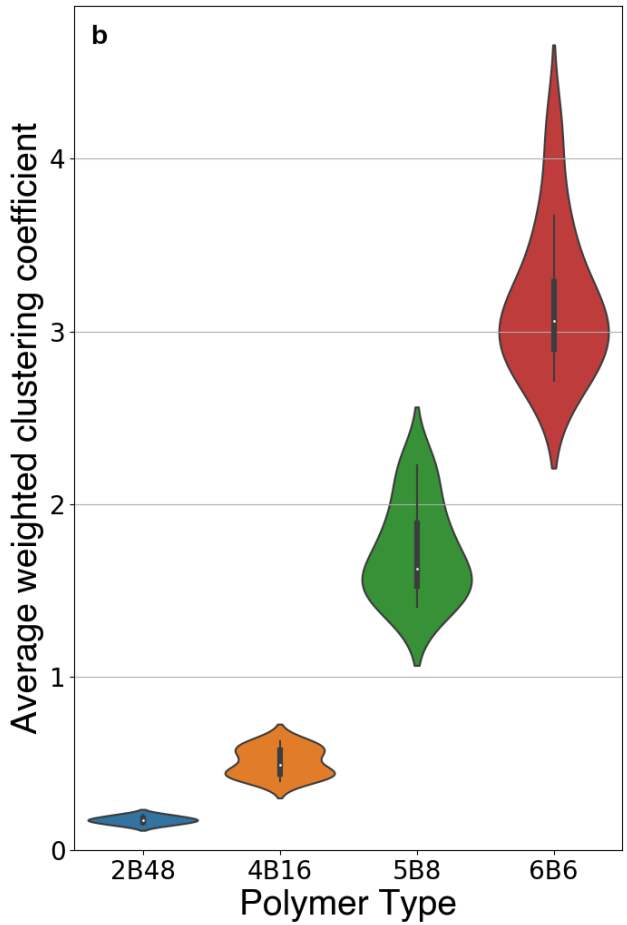
Weighted clustering coefficient

There is many definition for a weighted clustering coefficient, but they did not satisfy us, mainly because they tend to divide everything by the maximum weight of the edges.
Let $G(V, E)$ be a multigraph and A its adjacency matrix then the clustering coefficient of a node n_i is define as

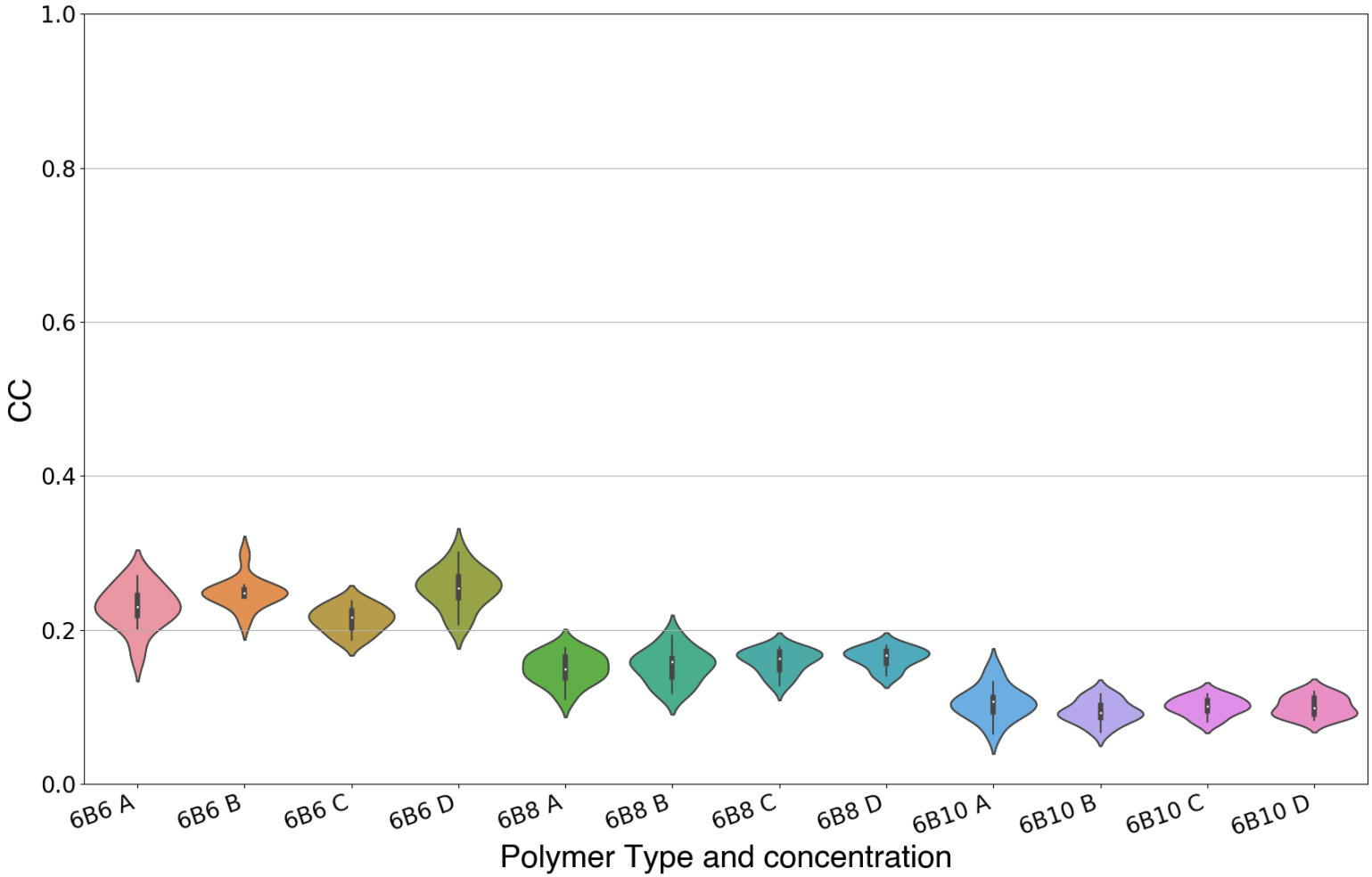
$$c_i^w = \frac{1}{k_i(k_i - 1)} \sum_{j,k} A_{i,j} A_{j,k} A_{k,i},$$

where k_i is the degree of n_i.
The clustering coefficient of G is $\frac{1}{|V|} \sum_i c_i^w$.

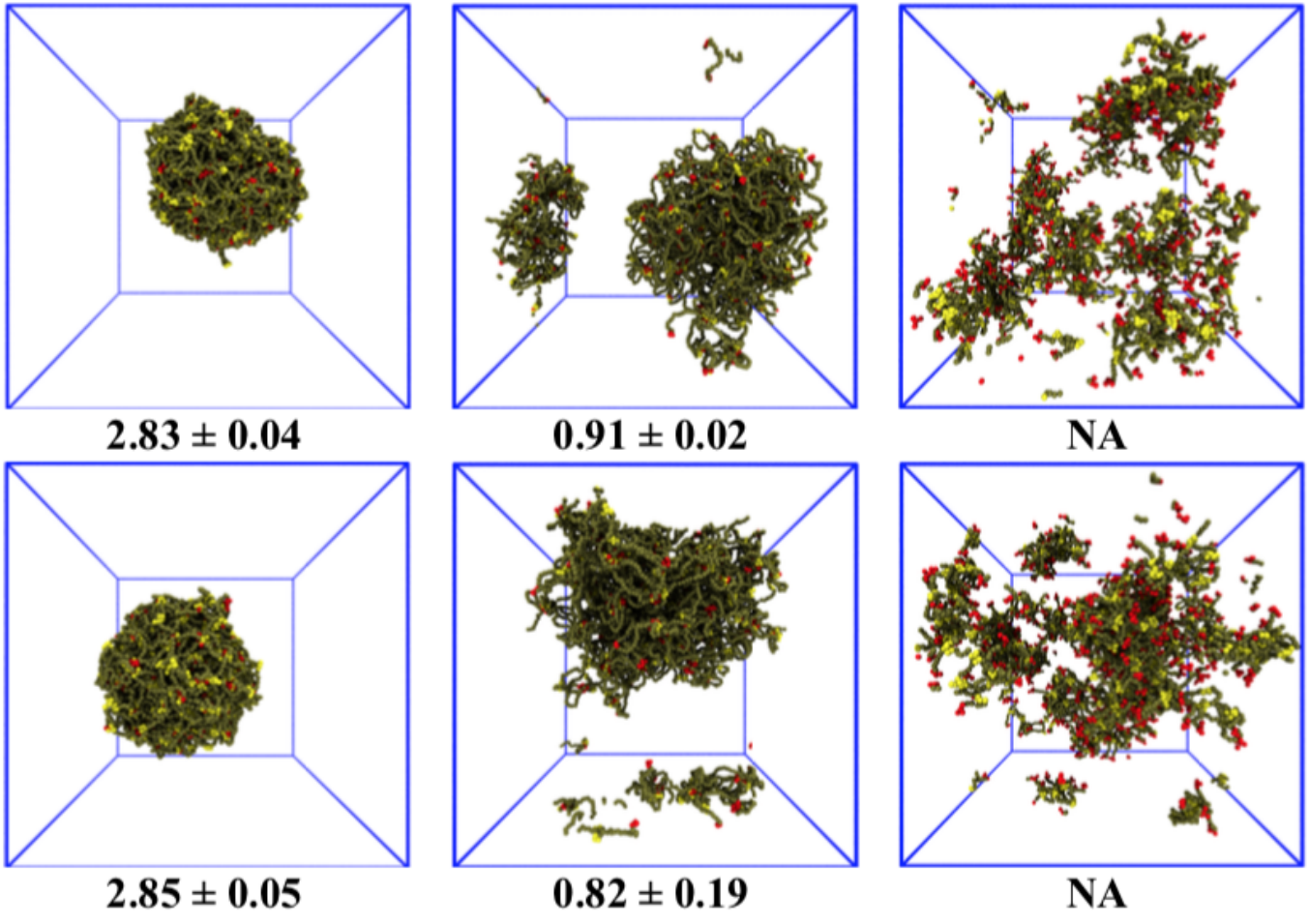
Results


b
Average weighted clustering coefficient
0
1
2
3
4
2B48
4B16
5B8
6B6
Polymer Type

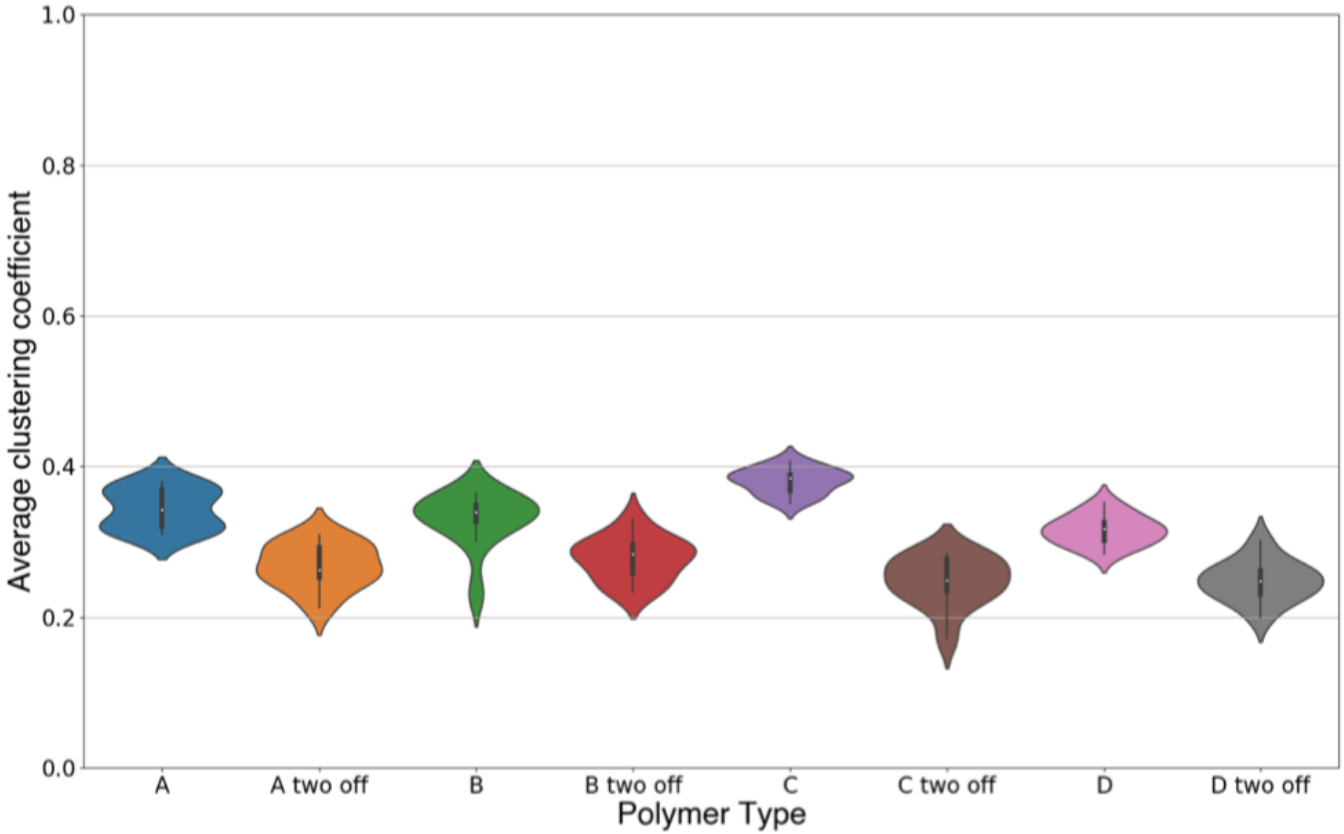
Varying the length of the chains ?



Disintegration



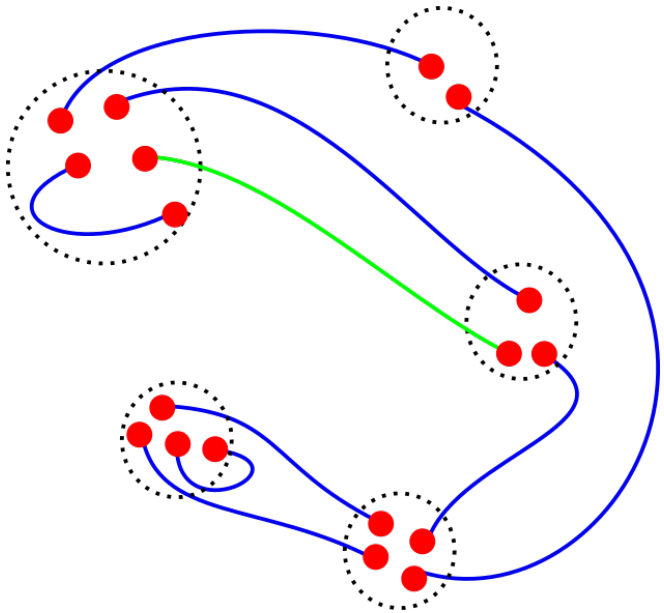
Disintegration



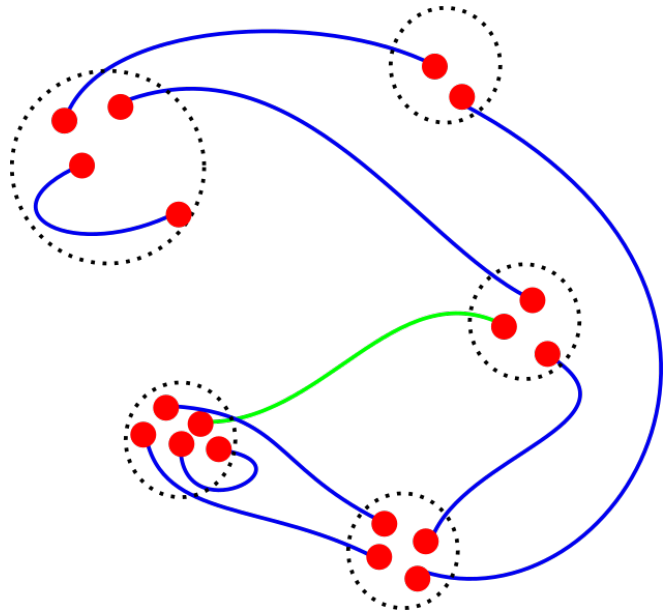
Motivation

We have many graphs of the same droplet corresponding to different times in the simulation and we want to compare them.
And it is really hard !
The graphs do not have the same numbers of vertices and edges so their adjacency matrix cannot tell us anything useful.
Moreover we have something that is multilevel : one for the polymer, one for the droplet.
In this work we concentrated on the first one.

Illustration



(a) $G(t)$



(b) $G(t+\tau)$

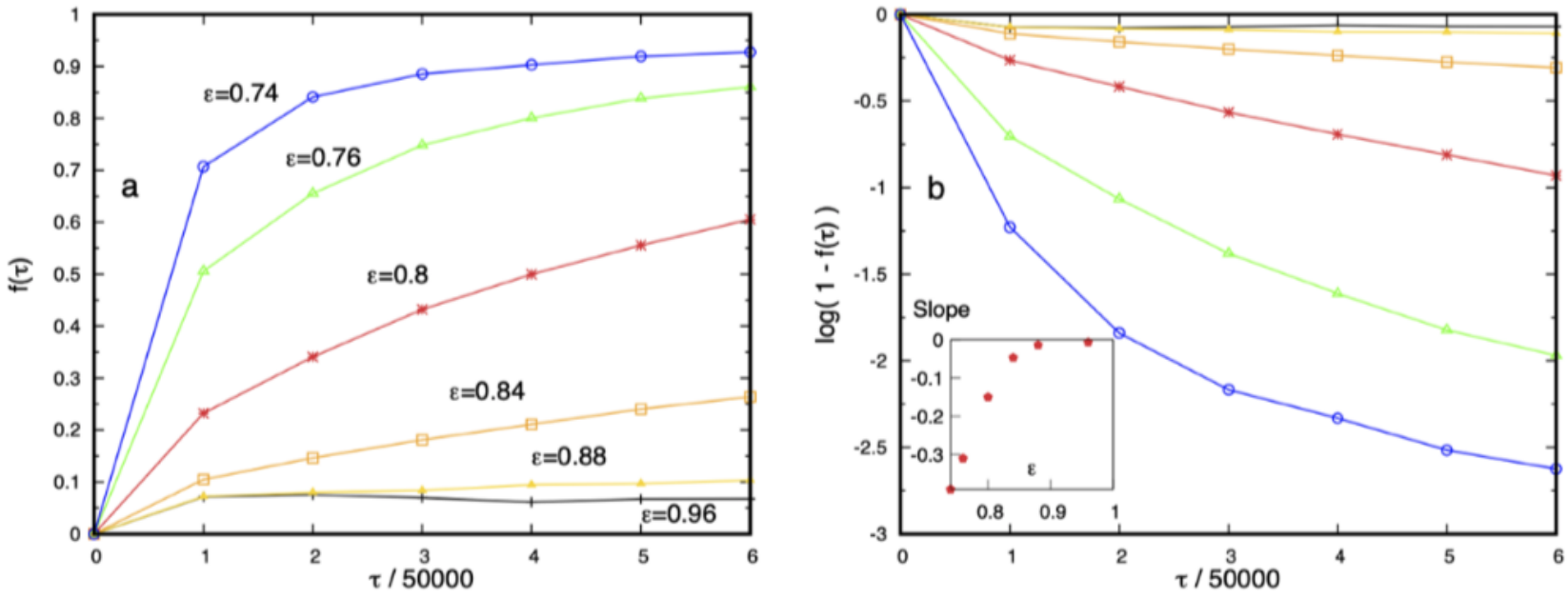
Fluidity

We define the fluidity of a network as :

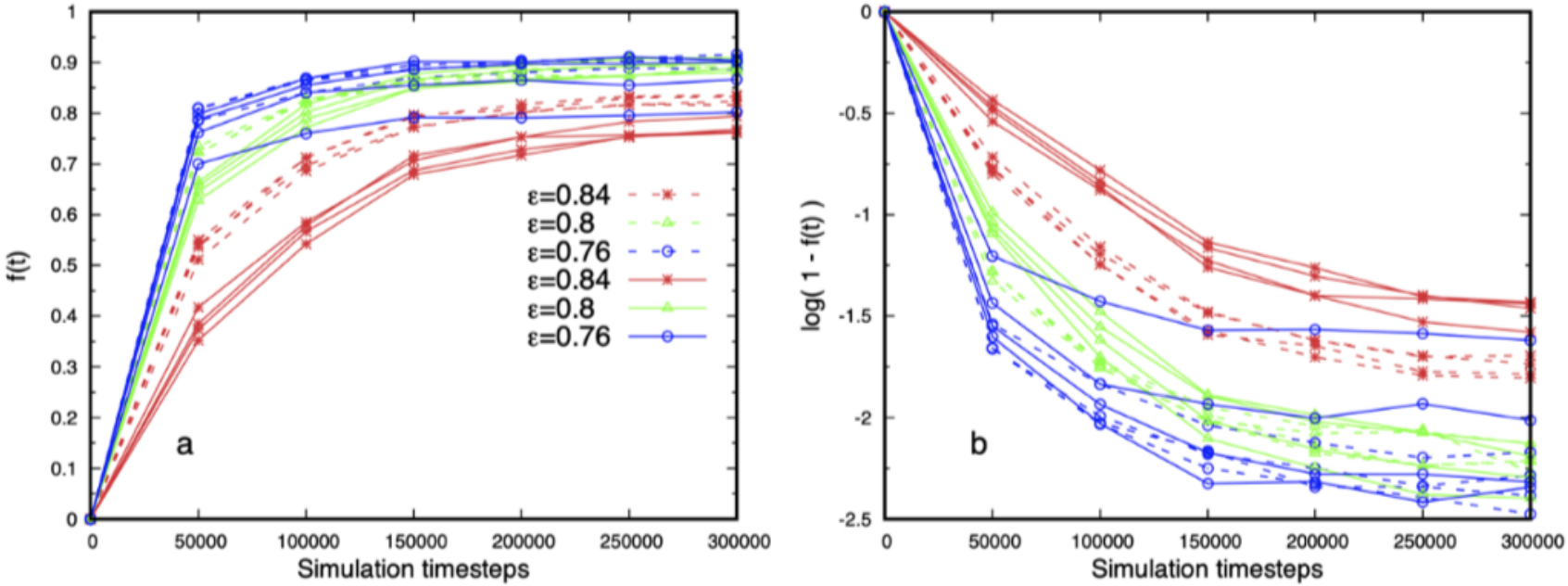
$$f(\tau) = \langle \frac{1}{N} \sum_i \frac{D_i(t, t+\tau) - 1}{M_i(t) - 1} \rangle,$$

where $\langle . \rangle$ signify an average on all the starting times t.

Results



Results


f(t)
Simulation timesteps
a
ε=0.84
ε=0.8
ε=0.76
ε=0.84
ε=0.8
ε=0.76
log(1 - f(t))
Simulation timesteps
b

Thank you

Thank you for your attention.