Characterisation of structures emerging from random colouring processes on a spatial graph

Silvia Rognone - Queen Mary University of London, School of Mathematical Sciences
Random and not-random
Random and not-random

Not Random
2020 USA elections

Random
20x20 sites grid
randomly coloured with 2 colours

11 May 2022 - Women in Mathematics Day
We consider a bi-dimensional squared lattice.

- We assigned to each site a colour with probability $p = \frac{1}{C}$.
- We call this process *random colouring process R*.
- Also in this simple case, we observe *structures*.
• We consider a bi-dimensional squared lattice.
• We assigned to each site a colour with probability \( p = \frac{1}{C} \)
• We call this process \textit{random colouring process R}.
• Also in this simple case, we observe \textit{structures}.
We investigate the probability distribution of $N_{\text{max}}$, i.e. the dimension of the largest structure observed in a single realisation of the process $R$.

- When $C$ decreases, the probability to observe large structures increases.

- The presence of structures is independent from the side $D$ of the lattice.
• A *cluster* is the maximal set of connected sites with the same colour.
• The *size* is the defined as the number of sites in this set and it is equal to $N$.
• The *frontier* of the cluster is the set of adjacent sites with a different colour.
• The *size of the frontier* is the defined as the number of sites in this set and it is equal to $f$. 

*Structural quantities*
**Structural quantities**

- **Surface**: the number $\sigma$ of edges shared between the cluster and the frontier.

- **Shape factor**: the ratio $S = \frac{N}{l^2}$
  - $S \in [0,1]$.
  - High $S$ is related to a dense structure.

- **Tree-likeness**: the ratio $\alpha = \frac{N - 1}{e_{in}}$
  - $\alpha \in (0.5,1]$.
  - High $\alpha$ is related to a large surface - small $\alpha$ to compact clusters.
Models

• At each time step, the frontier-sites are coloured either with the x colour, with probability \( p = \frac{1}{C} \), or with another colour, with probability \( q = 1 - p \).

• The seed cluster *grows limited in time* by the presence of other colours.

• At each time step, one of the frontier-sites is picked at random and coloured with the x colour with probability \( p = 1 \).

• The seed cluster *grows unlimited in time*.

11 May 2022 - Women in Mathematics Day
Results on structural quantities

- This is the trend for the sizes $N$ and $f$ of RGM clusters with $C=2$.
- Each $t$ represents a growth step.

- When $t=[0,20]$, $f$ exhibits a maximum.
- As expected, the size $N$ is comparable with the clusters size produced by the process $R$. 

11 May 2022 - Women in Mathematics Day
The peak in the distribution of shape factor for RGM clusters shifts to the left when $N$ increases, capturing their rarefied structure.

For EGM clusters, the shift is to the right, approaching the circle limit $S = 0.78$ for large $N$. 

11 May 2022 - Women in Mathematics Day
• RGM clusters present a high tree-likeness value, because of their surface-like nature.
• EGM clusters are reaching the lower limit of $\alpha = 0.5$ as expected for compact structures.
Time quantities

• The exit time $\tau_{i,F}$ from a site $i$ of the cluster $\mathcal{C}$ to one site of the frontier $F$, is defined as the number of steps a random walk needs to jump out of $\mathcal{C}$ starting from $i$. In the formula, $\pi_{ij}$ represents the probability to jump from the site $i$ to the site $j$.

$$\tau_{i,f} = 1 + \sum_{j \in \mathcal{C}, j \neq i} \pi_{ij} \tau_{j,f}$$

• The exit time $\tau$ from a cluster $\mathcal{C}$ is the average of the exit time over all the $N$ sites of $\mathcal{C}$. The mean exit time $\bar{\tau}$ is the average of the exit time over $r$ realisations of $\mathcal{C}$ at fixed $N$.

$$\tau = \frac{1}{N} \sum_{i}^{N} \tau_{i,f} \quad \bar{\tau} = \frac{1}{r} \sum_{l}^{r} \tau_{l}$$
Results on time quantities

- **Intuition**: the larger the surface, the less time spent inside a cluster.
- We propose a *Mean Field (MF) surface approximation* for $\tau$, as the ratio between $N_k$ and $\sigma$, where $k=4$ in the case of squared lattice.

\[
\tau_{MF} = \frac{N_k}{\sigma} \quad \bar{\tau}_{MF} = \frac{1}{r} \sum_{i} \tau_{MF,i}
\]
Results on time quantities

- EGM clusters exhibit a larger $\tau$ than RGM as a consequence of their compact structure. The mean exit time captures the structural differences between the two families of clusters.
- The MF surface approximation follows the data for RGM clusters but diverges for the EGM.
- $\bar{\tau}$ remains between the two theoretical limits $\bar{\tau}_Q$ and $\bar{\tau}_L$, i.e. the mean exit times from squares and lines, two families of clusters with respectively the lowest and the highest surface at fixed $N$.

11 May 2022 - Women in Mathematics Day
Application in real dataset

- Ongoing project on *plant roots cells dataset*.
- The spatial network is coloured with two colours. *Tree-likeness, size and time measures* are collected and analysed, compared to the random case.
- Roots with more *symmetrical spatial distribution* are more resilient to attacks.
Conclusions

• The *characterisation of random spatial structures* is crucial to *assign a statistical significance to the measured quantities*, when a colouring process is implemented on spatial networks.

• We show that *the mean exit time is an effective tool to classify these clusters* and to detect their structural properties.
Thanks for your attention!

s.rognone@qmul.ac.uk
@s_rognone

References: