



Research article

Towards a robust algorithm to determine topological domains from colocalization data

Alexander P. Moscalets^{1,2,3}, Leonid I. Nazarov⁴, and Mikhail V. Tamm^{1,4,*}

¹ Department of Applied Mathematics, National Research University Higher School of Economics, 101000, Moscow, Russia

² A.N. Nesmeyanov Institute of Organoelement Compounds of the Russian Academy of Sciences, 119991, Moscow, Russia

³ Chemistry Department, Moscow State University, 119991, Moscow, Russia

⁴ Physics Department, Moscow State University, 119991, Moscow, Russia

* **Correspondence:** Email: thumm.m@gmail.com; Tel: +79036103663;
Fax: +74959392988.

Abstract: One of the most important tasks in understanding the complex spatial organization of the genome consists in extracting information about this spatial organization, the function and structure of chromatin topological domains from existing experimental data, in particular, from genome colocalization (Hi-C) matrices. Here we present an algorithm allowing to reveal the underlying hierarchical domain structure of a polymer conformation from analyzing the modularity of colocalization matrices. We also test this algorithm on several model polymer structures: equilibrium globules, random fractal globules and regular fractal (Peano) conformations. We define what we call a spectrum of cluster borders, and show that these spectra behave strikingly differently for equilibrium and fractal conformations, allowing us to suggest an additional criterion to identify fractal polymer conformations.

Keywords: chromatin structure; fractal globule; Hi-C maps; complex networks; community detection

1. Introduction

The question of how an astonishingly long DNA chain consisting of 10^9 basepairs or even more folds into a compact state within a small volume of cell nucleus is still intriguing and not completely resolved. Indeed, not only is the packing compact which would have been easily achievable if DNA was in a so-called equilibrium globule state [1–3], but it is also capable to function biologically in a meaningful way. This biological function enforces many very specific and clearly non-equilibrium features including the existence of distinct chromosome territories and topological domains (TADs)

within single chromosome, easy unentanglement of chromosomes and chromosome parts (needed in preparation to mitosis, and during transcription), and ability of different parts of the genome to find each other in space strikingly fast in e.g. so-called promoter-enhancer interactions. The concrete mechanism which stabilizes these features is not yet completely understood. The main candidates for this stabilization is the so-called model of non-equilibrium fractal globule [4–14], as well as various models accentuating the formation of saturating bonds between the fragments of chromatin [15–22].

On the experimental side, the high-resolution data concerning the spatial organization of the genome is mostly due to the development of the genome-wide chromosome conformation capture (so called Hi-C) method [7, 23, 24] which allows to obtain the *colocalization* matrices of the genome packing, containing information about which particular genome fragments are *closed to each other* in space. It is from the statistical analysis of these matrices that the authors of [7, 8] deduced that the fractal globule state is a suitable candidate to describe chromosome packing. Originally, the Hi-C matrices represent the colocalization data averaged over many cells, however, recently a significant progress was reported in obtaining single-cell Hi-C maps [25].

Note therefore, that the data available include not the information about full spatial organization of the genome, but just on the parts of the genome which are spatially close to each other. It is, therefore, of great interest to develop methods extracting as much information as possible from such a dataset. In particular, it is a challenging question whether one can recover the information about the exact structure of TADs from colocalization data. It seems in principle reasonable to believe that such extraction is possible: indeed, parts of a chromosome belonging to a single TAD are spatially compact and should therefore more often find themselves in a close proximity with each other.

In this paper we suggest an algorithm, based on the methods of complex network theory [26–29] which allows to reveal a hierarchical TAD structure of a polymer conformation if it does exist, and check the applicability of this algorithm on several model polymer conformations. In what follows we discuss this algorithm for a particular case of a single-conformation colocalization matrix whose elements are ‘1’s and ‘0’s depending on whether the two corresponding monomers are spatially adjacent or not. The generalization for a more experimentally typical case of Hi-C maps *averaged over* many conformations is absolutely straightforward. Note, however, that the significance of resulting community structure is not ensured: indeed, as we argue elsewhere [12] the experimental Hi-C maps are in fact averaged over many substantially different folding conformations, and one expects therefore the community structure of the average to be significantly less rich than the community structure of the individual conformations. We hope, however, that (i) further advances in the single-cell Hi-C mapping techniques will allow in-depth analysis of the community structure of individual genome conformations, and (ii) that comparison of the community structures corresponding to individual and averaged Hi-C maps may shed light on which characteristics of genome folding are conserved from realization to realization and which ones are variable.

2. Model Description

It is feasible to construct a mapping between any given configuration of a polymer chain of N monomers and a colocalization matrix \mathbf{W} with matrix elements w_{ij} equal to 1 if i -th and j -th monomers are close to each other in space (i.e., if the spatial distance r_{ij} is less than some cut-off value a which here and below is presumed to be of order of the monomer-monomer bond length) and 0 otherwise. Our

aim is to extract the information about topological domain structure of the chain conformation from this matrix \mathbf{W} , and to estimate if this domain structure is stable (in some sense to be determined later), and if it possesses ultrametric properties (i.e., consists of sequentially nested domains of smaller and smaller size) expected from the structure of fractal globule (see, e.g., [4, 8, 12, 14] for the discussion of ultrametricity in the context of fractal globules). In what follows we develop a method allowing us to do that and show the results obtained by this method on several test polymer configurations. Application of this method to real experimental data goes beyond the scope of this paper and will be provided elsewhere.

The analysis of matrix \mathbf{W} starts from the notion that it could be reinterpreted as an adjacency matrix of some complex graph (network), which allows us to use the community detection methods developed in the complex network theory [30]. Indeed, it seems natural to assume that monomers belonging to the same TAD will more often be in spatial proximity to each other in space than monomers belonging to different TADs, and therefore a community of comparatively well-connected nodes in the network theory sense can be a good proxy to a real topological domain. It is known, however, that the problem of optimal division of a network into a set of communities (comparatively well connected clusters) is ill-posed and does not have a single definite answer; there exist numerous techniques based on spectral properties of adjacency matrix [31], non-backtracking matrix [32], synchronization in networks [33]. In what follows we employ the modified modularity-optimization method as developed in [26–29] which we find most appropriate to our needs as it very naturally allows to look into the community structure of networks on different scales.

Consider a partitioning of an N -node network into a set of k clusters (see Fig.1 below for a toy example of such partitioning). Any such partitioning can be described by a matrix \mathbf{C} of size $N \times k$ with $c_{i\alpha} = 1$ if i -th element of the network belongs to a α -th cluster and 0 otherwise. Naturally, one assumes $\sum_{\alpha} c_{i\alpha} = 1$ for any i (i.e., each node belongs to one and only one cluster). Then, according to [26] the modularity $Q[\mathbf{W}, \mathbf{C}]$ of such a partition is defined as

$$Q[\mathbf{W}, \mathbf{C}] = \frac{1}{2w} \sum_{i,j} \left(w_{ij} - \frac{w_i w_j}{2w} \right) \sum_{\alpha} c_{i\alpha} c_{j\alpha}, \quad (1)$$

where w_{ij} are the elements of \mathbf{W} , $w_i = \sum_j w_{ij}$ is the total number of neighbors of i -th node (also called ‘strength’ of the node), $w = \sum_{ij} w_{ij}/2$ is the total number of links (strength) of the network; the sum over α in (1) equals 1 if i -th and j -th monomers belong to the same cluster, and 0 otherwise. The original modularity-based community detection algorithm demands to maximize the functional Q with respect to k and \mathbf{C} , the corresponding maximizing separation is then considered to be optimal. Thus defined method of modularity optimization is known to have two important drawbacks. First, it is known to have a so-called resolution limit, so that it is impossible to find any clusters of size less than $\sim \sqrt{2w}$ correctly [27]. Second, the total number of possible partitions of a network into clusters is exponentially large in N making the search of an optimal partition an NP-complete problem (i.e., a problem whose shortest possible solution time grows exponentially with N , see [34, 35] for the introduction to this concept).

There exists, however, a possibility to circumvent the first problem. Following [28, 29] introduce the so-called resistance parameter r , i.e., introduce a modified adjacency matrix $\bar{\mathbf{W}}(r)$ whose matrix elements equal to $\bar{w}_{ij} = w_{ij}$ for $i \neq j$, and $\bar{w}_{ii} = r$. In terms of the underlying network this corresponds to adding self-loops with weight (strength) r , which is, generally speaking, non-integer and can be even

negative. Now, proceed with the optimization of the modularity functional (1) for this new matrix $\bar{\mathbf{W}}(r)$ (note that the definition of modularity never relies on the elements of adjacency matrix being Boolean variables, indeed, it was originally introduced for weighted networks). The larger the resistance parameter r , the more are nodes coupled to themselves as compared to other surrounding nodes. As a result of that, as shown in [29] smaller and smaller clusters get determined. Indeed, if r is larger than some network-dependent critical value r_{\max} the optimal partition is one separating the network into N clusters consisting of 1 monomer each. On the other hand side, if r is less than some (once again, network dependent) r_{\min} which is usually negative, the optimal partition consists of a single cluster covering the whole network.

As for the second drawback of usual modularity algorithm, i.e. the exponential increase of the possible number of separations with the size of the system, it also can be circumvented in this particular case. Indeed, remind that we originally defined the adjacency matrix \mathbf{W} as a colocalization matrix of some *polymer* configuration. That is to say, the monomers of the network are naturally numbered along the chain, and monomers close along the chain are automatically close in space due to the connectivity of a polymer. This allows us to postulate by definition that we only consider clustering partitions which separate a chain in fragments which are adjacent along the chain, i.e. if i -th and j -th monomer belong to the same cluster, than any k -th monomer with $i < k < j$ belongs to the same cluster as well. Note, that such definition of clusters is in accordance with how topological domains are usually understood in polymer and biophysical literature: parts of the chain that are close *both* along the chain and in real space. Simultaneously, it is easy to see that such a restriction on possible partitioning reduces their overall number from exponential in N to quadratic in N , allowing us to produce a rather fast deterministic partitioning algorithm, which we realized in FORTRAN 95.

For the purposes of our work, it is instructive to consider how the resulting partitioning (community structure) *evolves* with the change of r . Indeed, if one considers a completely random Erdős–Rényi network, one would expect that at r_{\min} the network is separated into two clusters of roughly same size, then as the value of resistance reaches some $r_2 > r_{\min}$ it separates into three clusters of, once again, roughly same size, then at $r_3 > r_2$ into four clusters, etc. The important thing is that in the absence of any underlying structure of the network one expects the cluster boundaries to be essentially uncorrelated: every time the number of clusters increases by one and *all* cluster boundaries rearrange. On the other hand, if the networks has an underlying structure of hierarchically organized TADs, one expects the increase of r to cause not a complete rearrangement of cluster structure, but a decomposition of already existing clusters into smaller parts with boundaries of larger clusters remaining essentially stable. In order to quantify this qualitative notion we introduce here the concept of a *spectrum of borders*.

Let $H_i(r)$ ($i = 1, \dots, N - 1$) be a Heaviside step function indicating whether the bond between i -th and $(i + 1)$ -th monomer is a border of a cluster at given r (that is to say $H_i(r) = 1$ if the said bond is a boundary, and $H_i(r) = 0$ otherwise). Then the total fraction of the range (r_{\min}, r_{\max}) when a given bond is a border of a cluster is given by

$$f_i = (r_{\max} - r_{\min})^{-1} \int_{r_{\min}}^{r_{\max}} H_i(r) dr. \quad (2)$$

The higher f_i is, the more stable is the cluster border at $(i, i + 1)$ bond. The whole set of f_i for all i is what we call a spectrum of borders (SoB). Inspecting such a spectrum one should be able to estimate

how stable the cluster structure of the network is, and what exactly is the natural hierarchical domain structure of a network, if any. If there is no well-defined clusters in a network, the SoB is a more or less uniform distribution of lines, while if there is a small number of f_i significantly exceeding the rest, it signifies a network with stable and well-defined cluster structure.

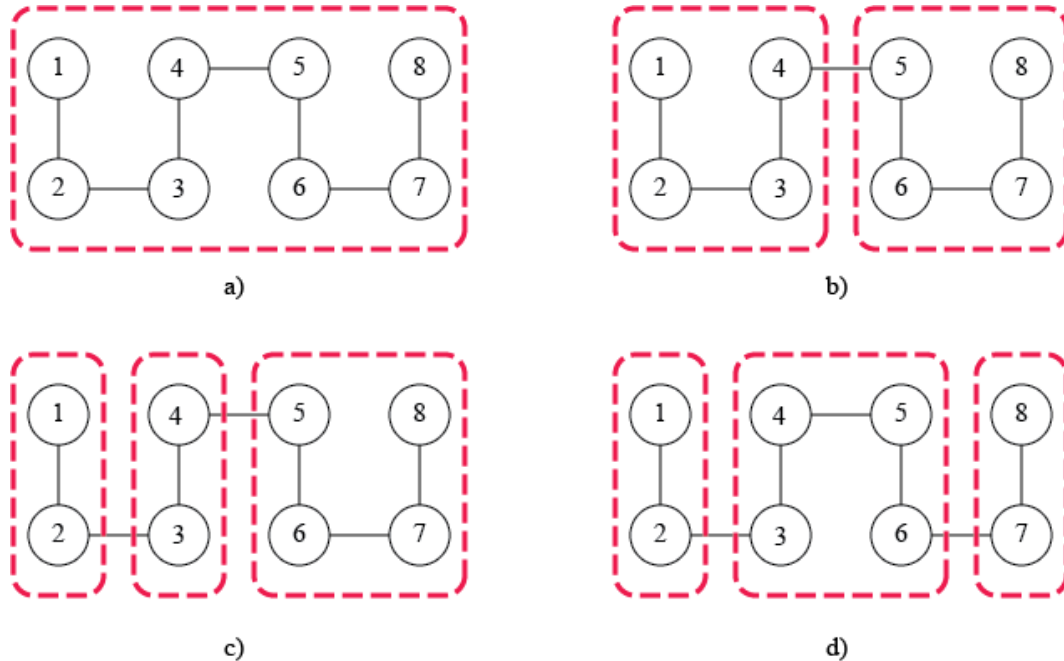


Figure 1. A toy model of a network consisting of 8 nodes and its possible separations into clusters. See explanation in the main text for details.

To clarify this definition, let us start with a toy example. Consider a model network with 8 sequentially numbered nodes (see Fig.1a) which we study in the interval between some fixed r_{\min} and r_{\max} . Assume first that along the whole interval it is partitioned into exactly two clusters with the border at node 4 (Fig.1b). Then

$$f_i = \begin{cases} 1, & \text{for } i = 4; \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

This means that in the whole range of r there is only one border between two clusters and this border does not shift to the left or right, thus the clusters always contain exactly 4 nodes each (i.e., if partitioning shown in Fig.1b holds for any r , then the SoB is the one shown in Fig.2a). If there appears a region within which these clusters vanish and optimal partitioning consists of just one cluster (i.e., Fig.1b for some r s and Fig.1a for some others), then the amplitude f_4 becomes less than 1 (Fig.2b).

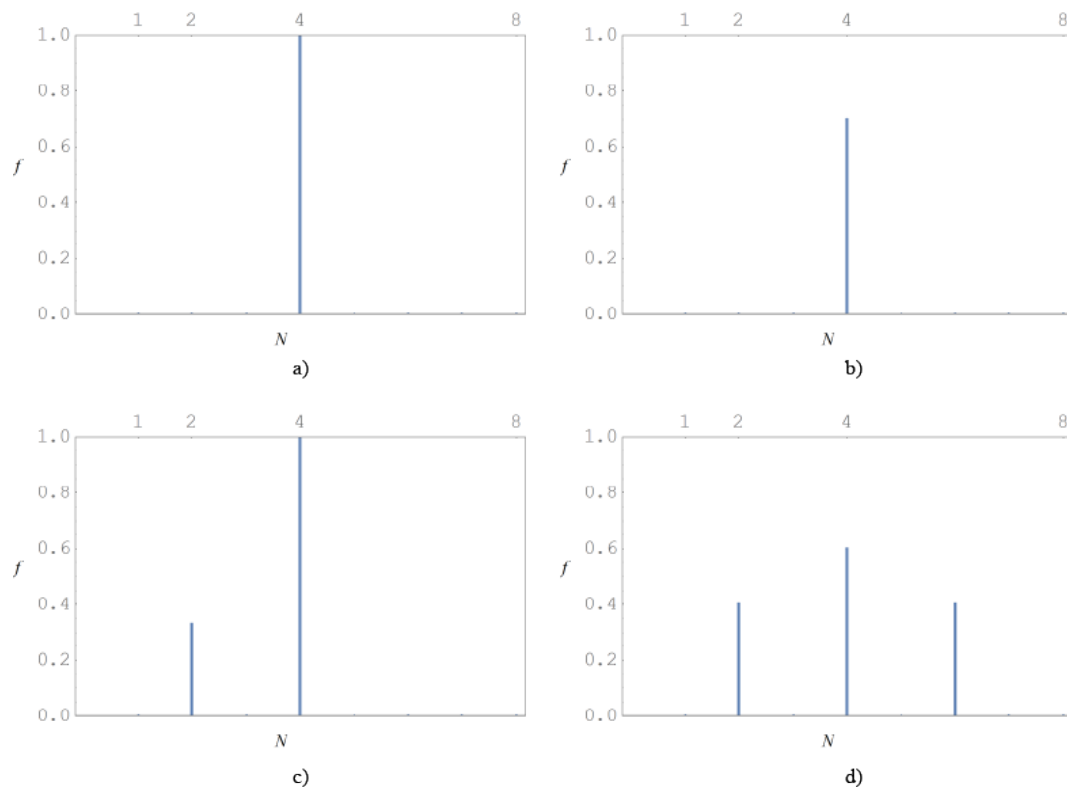


Figure 2. Spectra of borders for a toy eight-nodes network shown in Fig.1. For the explanation of differences between the boxes see the main text.

Suppose now, that above certain value $r_2 \in (r_{\min}, r_{\max})$ the first of the two clusters splits into 2 smaller ones of size 2 each (see Fig.1c). This will result in the emergence of new line of smaller amplitude in the SoB at node 2 (Fig.2c):

$$f_i = \begin{cases} 1, & \text{for } i = 4; \\ f, & \text{for } i = 2; \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where

$$f = \frac{r_{\max} - r_2}{r_{\max} - r_{\min}}, \quad 0 < f < 1. \quad (5)$$

A different possibility is that at $r_2 \in (r_{\min}, r_{\max})$ the border at node 4 disappears and new set of three clusters consisting of, say, nodes (1, 2), (3, 4, 5, 6), and (7, 8) arise (see Fig.1d). Then the set of amplitudes in the SoB will read

$$f_i = \begin{cases} 1 - f, & \text{for } i = 4; \\ f, & \text{for } i = 2, 6; \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

The corresponding spectra of borders are shown in Fig.2c and Fig.2d, respectively.

Thus, in accordance with what have been said above, if the borders of clusters dangle more or less randomly along the chain, the resulting SoB consists of a number of low peaks of roughly equal height, and such a behavior is characteristic for networks with labile, fuzzy domain structure. On the contrary, the changes when the existing large cluster splits into smaller ones without changing its outer borders give rise to a series of peaks which significantly differ in height. Such behavior seems to be characteristic of networks with well-defined hierarchical domain structure.

3. Results and Discussion

To check the aptitude of the described approach we applied it to series of adjacency matrices of several model polymer conformations. In particular, we studied, (i) a completely deterministic Peano curve which is the simplest possible proxy of a fractal globule conformation with most prominent hierarchical self-similar cluster organization possible, (ii) an equilibrium conformation of a Gaussian polymer globule, (iii) a random fractal globule conformation obtained by the conformation-dependent polymerization.

The first two conformations are more or less standard, while the third, suggested originally in [13] (the algorithm is briefly outlined below, and we address the reader to the supplementary materials of that paper for full details about this algorithm) is, in our opinion, one of the best existing candidates to represent the generic metastable fractal globule state. It shows significant stability in dynamic computer simulations and its statistical characteristics are very similar to what is obtained in other fractal-globule generating algorithms. Therefore, the main aim of our test is to show that random fractal globule conformations could be robustly and reproducibly distinguished from equilibrium Gaussian ones based on their SoBs.

The length of polymer under consideration was $N = 4096$ for Peano curve and $N = 10000$ for random fractal and equilibrium conformations. All chain configurations were generated on cubic lattice. For generation of the fractal globule the following rules were used. The monomers are added sequentially to the chain, with a new monomer added to one of the 6 nodes of a lattice adjacent to the endchain monomer with probabilities p_i , $i = 1, \dots, 6$ to go in each direction equal to:

$$p_i = \mathcal{N}^{-1} \begin{cases} 10^{-6}, & \text{if } i\text{-th site is visited,} \\ 1 + 10^4 n_i, & \text{if } i\text{-th site is not visited,} \end{cases} \quad (7)$$

where $\mathcal{N} = \sum_i p_i$ is a normalizing constant, and n_i is the number of occupied sites within a unit sphere centered at site i . The significant difference with [13] is that the fractal globules were obtained in a free volume without periodic boundary conditions, which allowed to obtain conformations with very developed surface. Equilibrium globule was obtained as random walk within a sphere of radius $R = 1.6 \left(\frac{3N}{4\pi}\right)^{1/3}$ and reflecting surface. In case of trapped configurations the chain end can go back through randomly chosen, among already visited, site.

The elements of adjacency matrix w_{ij} equal to 1 (monomers are considered to be neighbors) if they are separated by distance $r_{ij} \leq a = \sqrt{3}$ lattice units. In our analysis we somewhat arbitrarily have chosen $r_{\min} = -20$ and $r_{\max} = 80$. While $r_{\min} = -20$ roughly corresponds to the point where the first partitioning of a network into two clusters takes place for the networks under study, the chosen value of $r_{\max} = 80$ is significantly smaller than the natural limit defined above. That is to say, even for largest r under consideration, our networks are quite far from being separated into N single node clusters. This

choice was dictated mostly by the saving of CPU modeling time. The calculations presented below thus grasp the essential behavior of the SoB for large clusters and the most stable domain walls, which are, in our opinion, of the most interest. Indeed, small-scale clustering structure of our test configurations can be significantly plagued by the underlying discrete lattice. However, we perform a check of the robustness of suggested algorithm with respect to the choice of r_{\max} (see below).

The application of the proposed algorithm to Peano curve reveals hierarchically nested clustering, as expected. At values $r < -20.7$ there is the only community with 4096 nodes. At $r \approx -20.7$ two clusters, consisting of 2048 nodes each, appear. With further increase of r each of these clusters splits exactly into 2 half-size clusters, but the border at node $i = 2048$ remains at place, then each of the four clusters splits into halves, etc, all the way until $r = r_{\max}$. Therefore, f_{2048} is the highest peak in SoB for Peano curve, followed by f_{1024} and f_{3072} , etc. We get finally a SoB consisting of an hierarchical set of equidistant peaks as shown in Fig.3. Note that smaller clusters positioned deep inside the globule and on its outer surface behave somewhat differently, which explains why starting from the third generation the values of peaks of the same generation are not exactly equal.

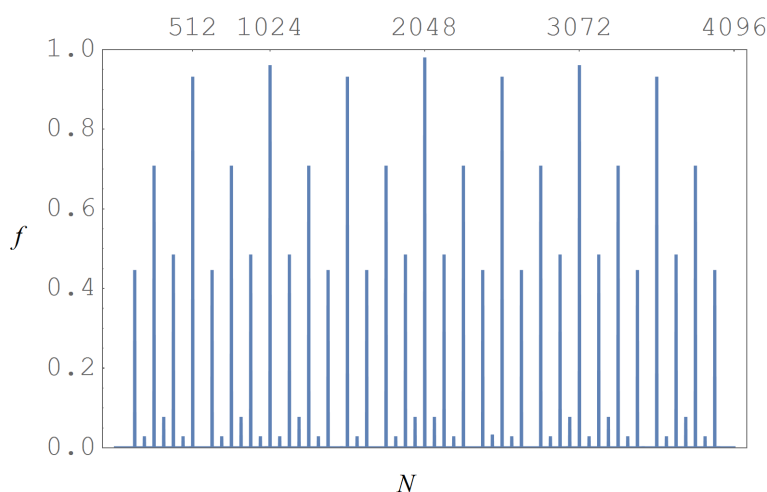


Figure 3. Spectrum of borders (magnitude f vs node number N) obtained for complex network corresponding to Peano curve.

The Fig.3 shows six levels of hierarchical organization of Peano curve, each larger cluster (consisting of the nodes between two largest peaks) having a fully deterministic internal structure. For example, the second level cluster, consisting of the nodes 2048 to 3072, has two domains (divided by line at node 2560) of 512 nodes each. The fourth level of hierarchy is organized by eight peaks on nodes $256 + 512n$ ($n = 0, \dots, 7$) with those of higher level, which make it a total of 16 clusters. Such a picture of discrete lines with gradually lowering heights at each level is a qualitative representation of regularly fractal, self-similar polymer globule.

Consider now a random fractal globule. Clearly, it does not have such a regular fully symmetric domain structure (see Fig.4), whatever one may notice that the peaks in the structure are once again very widely distributed: there are some relatively very high peaks (e.g., ones at nodes 4441, 9849, 2844 reach well above 0.8) and the general structure of the SoB is very widely disperse.

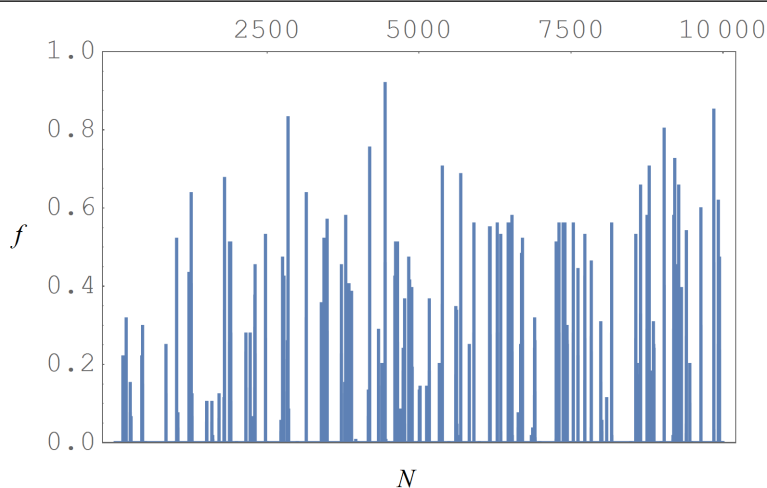


Figure 4. Spectrum of borders (magnitude f vs node number N) obtained for complex network corresponding to fractal (sticky) globule.

Contrary to that, the SoB for an equilibrium globule (Fig.5) is essentially a dense forest of peaks, peaks of any height seem to appear with equal probability without any gaps between higher and lower peaks. Such a picture tells us that almost each of the node has been the border of any cluster, or in other words, as it was described above, on magnifying the network the cluster structure rearranges completely with changing r , confirming thus that an equilibrium globule does not have a well-defined cluster structure.

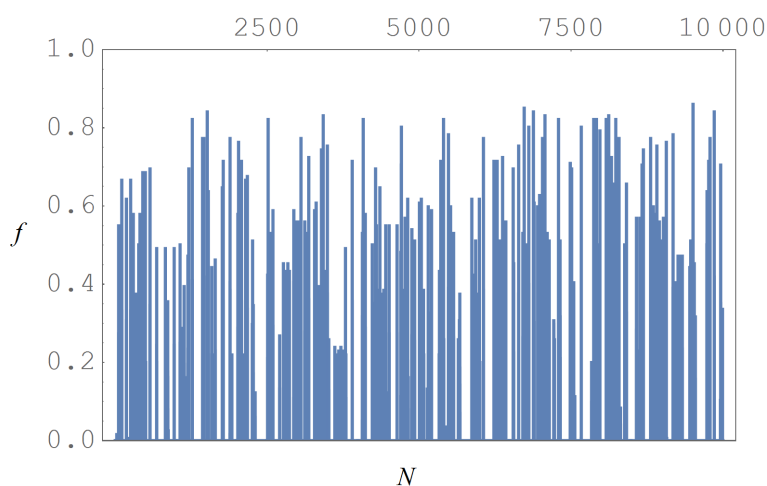


Figure 5. Spectrum of borders (magnitude f vs node number N) obtained for complex network corresponding to equilibrium globule.

In order to make the difference between figures Fig.4 and Fig.5 more clear, it is instructive to reorder peaks of the SoB in descending order, constructing thus an *ordered SoB*, where the highest peak is renamed f_1 , the second highest f_2 , and so on. Fig.6 shows the envelope lines of ordered SoBs

for 10 different random fractal realizations (all of them cluster in the right bottom of the picture) and 10 different equilibrium globule conformations (which cluster in the top left). One sees immediately that spectra of fractal and equilibrium globules are clearly distinguishable: spectra of equilibrium conformations show approximately linear descent, confirming that in such conformation borders of all intensities occur with roughly the same probability, while fractal globule has a clear cluster structure with a small number of very strong borders dominating over the rest.

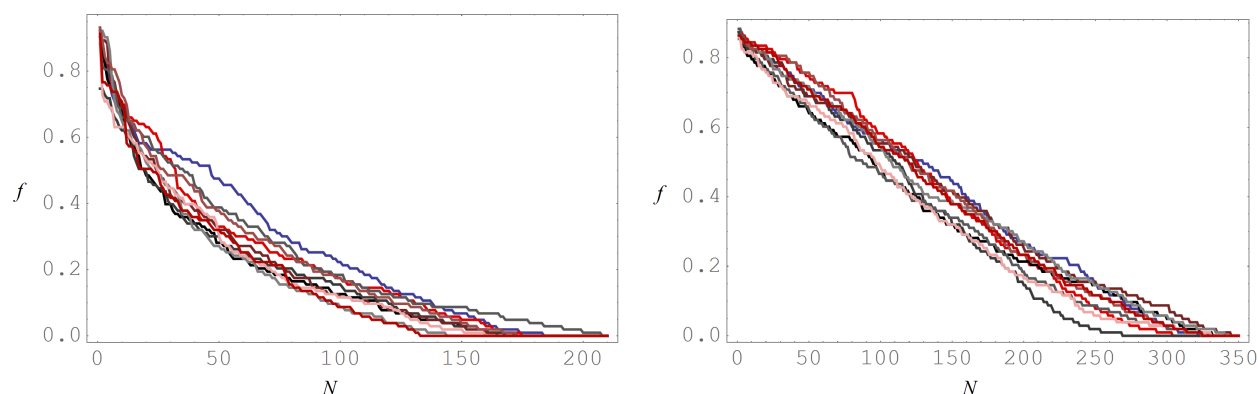


Figure 6. Ordered amplitudes f from spectrum of borders for 10 different realizations of fractal (left) and equilibrium (right) globules.

Moreover, in order to check if this distinction between two classes of conformations is robust with respect to change in r_{\max} , we have plotted a series of ordered SoBs for a single random fractal and a single equilibrium conformation but different values of r_{\max} varying from 80 to 140, see Fig.7. It is clear from the Figure that the changes of the curves, although clearly visible, do not change the general result: once again the curves for fractal and equilibrium conformations are clearly distinguishable.

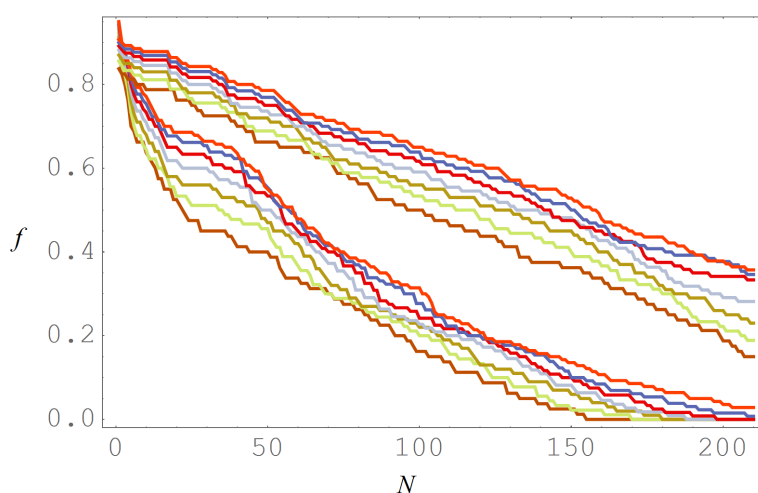


Figure 7. Comparison of ordered amplitudes f of fractal (bottom) and equilibrium (top) globules at different values of parameter r_{\max} chosen from interval 80 to 140. Curves of the same color are obtained at the same values of r_{\max} .

Fig.8 shows the average shape of ordered SoBs, obtained from averaging the curves shown in Fig.6 over 10 different conformations of the same class. Clearly, the spectra have very different shape, the curve corresponding to the equilibrium globule can be fitted as a straight line, whereas the case of fractal globules highly non-linear with a steep descent at the beginning. We are now working on a possibility to construct theoretical explanations for the shapes of the obtained curves, corresponding results will be presented elsewhere.

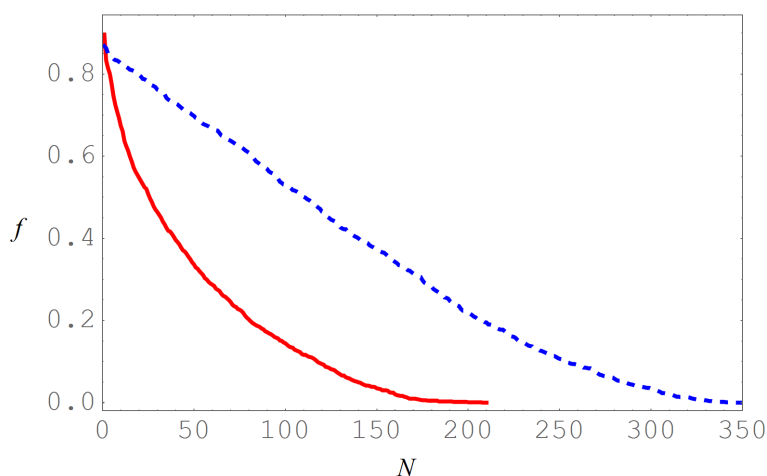


Figure 8. Comparison of averaged ordered amplitudes f of fractal (red solid) and equilibrium (blue dashed) globules.

We expect that the concept of the ordered spectrum of borders introduced above will provide an additional tool to differentiate between different possible conformation types of the polymer molecules, while the exact positions of the most strong borders will provide information about the position of TAD borders in real chromosome conformations.

4. Conclusion

In conclusion, let us discuss once again the possible applicability of the algorithm suggested in this paper to the experimental Hi-C maps, which are usually averaged over millions of cells with, generally speaking, different folding structures. Such maps are in fact symmetric matrices with non-negative non-integer elements, and they can be naturally interpreted as adjacency matrices of *weighted* complex networks. The algorithm we suggest does not at any point rely on the fact that the network under consideration is unweighted, and thus can be applied without change to these averaged Hi-C maps. There is, however, a more subtle question. It is not completely clear how much of the original hierarchy of the chain packing is conserved from realization to realization. We expect, and plan to check elsewhere, that those parts of the TAD structure which are repeated in all (or most) cells, should be accessible from analysing community structure of averaged Hi-C maps. On the contrary, those parts which are different from cell to cell should be observable only from the single-cell Hi-C maps (provided the corresponding experimental techniques will progress) but we expect them to be smeared over in the averaged maps. We expect that the comparison of the community structures of the averaged

and single-cell Hi-C maps can elucidate (especially coupled with the progress of the single-cell Hi-C mapping techniques) the question of how stable individual TADs are from cell to cell (see, e.g. [12] for further discussion of this subject).

Acknowledgments

Authors are grateful to S. K. Nechaev, L. Mirny, A. V. Chertovich, and P. Kos for fruitful discussions and to A. Cherstvy for his useful comments on the text of the manuscript. The work is partially supported by the RFBR grant 14-03-00825 and by the Higher School of Economy program for basic research.

Conflict of Interest

All authors declare no conflicts of interest in this paper.

References

1. de Gennes PG (1979) *Scaling Concepts in Polymer Physics*, Cornell University Press.
2. Grosberg AY, Khokhlov AR (1994) *Statistical Physics of Macromolecules*, AIP Press.
3. Rubinstein M, Colby RH (2003) *Polymer Physics*, Oxford:OUP.
4. Grosberg AY, Nechaev SK, Shakhnovich EI (1988) The role of topological constraints in the kinetics of collapse of macromolecules. *J Phys France* 49(12):2095–2100.
5. Grosberg A, Rabin Y, Havlin S, et al. (1993) Crumpled globule model of the three-dimensional structure of DNA. *EPL-Europhys Lett* 23(5):373.
6. Rosa A, Everaers R (2008) Structure and dynamics of interphase chromosomes. *PLoS Comput Biol* 4(8):e1000153.
7. Lieberman-Aiden E, van Berkum NL, Williams L, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950):289–293.
8. Mirny LA (2011) The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res* 19(1):37–51.
9. Halverson JD, Smrek J, Kremer K, et al. (2014) From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Rep Prog Phys* 77(2):022601.
10. Grosberg AY (2014) Annealed lattice animal model and Flory theory for the melt of non-concatenated rings: towards the physics of crumpling. *Soft Matter* 10:560–565.
11. Rosa A, Everaers R (2014) Ring polymers in the melt state: The physics of crumpling. *Phys Rev Lett* 112:118302.
12. Nazarov LI, Tamm MV, Avetisov VA, et al. (2015) A statistical model of intra-chromosome contact maps. *Soft Matter* 11(5):1019–1025.
13. Tamm MV, Nazarov LI, Gavrilov AA, et al. (2015) Anomalous diffusion in fractal globules. *Phys Rev Lett* 114:178102.

14. Bunin G, Kardar M (2015) Coalescence model for crumpled globules formed in polymer collapse. *Phys Rev Lett* 115:088303.
15. Sachs RK, van den Engh G, Trask B, et al. (1995) A random-walk/giant-loop model for interphase chromosomes. *P Natl Acad Sci* 92(7):2710–2714.
16. Münkler C, Langowski J (1998) Chromosome structure predicted by a polymer model. *Phys Rev E* 57:5888–5896.
17. Ostashevsky J (1998) A polymer model for the structural organization of chromatin loops and minibands in interphase chromosomes. *Mol Biol Cell* 9(11):3031–3040.
18. Mateos-Langerak J, Bohn M, de Leeuw W, et al. (2009) Spatially confined folding of chromatin in the interphase nucleus. *P Natl Acad Sci* 106(10):3812–3817.
19. Iyer BVS, Arya G (2012) Lattice animal model of chromosome organization. *Phys Rev E* 86:011911.
20. Barbieri M, Chotalia M, Fraser J, et al. (2012) Complexity of chromatin folding is captured by the strings and binders switch model. *P Natl Acad Sci* 109(40):16173–16178.
21. Fritsch CC, Langowski J (2010) Anomalous diffusion in the interphase cell nucleus: The effect of spatial correlations of chromatin. *J Chem Phys* 133(2):025101.
22. Fritsch CC, Langowski J (2011) Chromosome dynamics, molecular crowding, and diffusion in the interphase cell nucleus: a monte carlo lattice simulation study. *Chromosome Res* 19(1):63–81.
23. Dekker J, Rippe K, Dekker M, et al. (2002) Capturing chromosome conformation. *Science* 295(5558):1306–1311.
24. Naumova N, Imakaev M, Fudenberg G, et al. (2013) Organization of the mitotic chromosome. *Science* 342(6161):948–953.
25. Nagano T, Lubling Y, Stevens TJ, et al. (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502(7469):59–64.
26. Newman MEJ (2004) Analysis of weighted networks. *Phys Rev E* 70:056131.
27. Lancichinetti A, Fortunato S (2011) Limits of modularity maximization in community detection. *Phys Rev E* 84:066122.
28. Granell C, Gomez S, Arenas A (2012) Hierarchical multiresolution method to overcome the resolution limit in complex networks. *Int J Bifurcat Chaos* 22(07):1250171.
29. Arenas A, Fernandez A, Gomez S (2008) Analysis of the structure of complex networks at different resolution levels. *New J Phys* 10(5):053039.
30. Fortunato S (2010) Community detection in graphs. *Phys Rep* 486(3-5):75–174.
31. Newman MEJ (2006) Modularity and community structure in networks. *P Natl Acad Sci* 103(23):8577–8582.
32. Krzakala F, Moore C, Mossel E, et al. (2013) Spectral redemption in clustering sparse networks. *P Natl Acad Sci* 110(52):20935–20940.
33. Arenas A, Diaz-Guilera A, Kurths J, et al. (2008) Synchronization in complex networks. *Phys Rep* 469(3):93–153.

-
34. Sethna JP (2006) *Statistical Mechanics: Entropy, Order Parameters and Complexity*, Oxford:OUP.
35. Mezard M, Montanari A, (2009) *Information, Physics, and Computation*, Oxford:OUP.



©2015, Mikhail V. Tamm, et al., licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)