SCIENTIFIC REP⚙RTS

OPEN

Hidden geometries in networks arising from cooperative self-assembly

Milovan Šuvakov [iD] [1,2], Miroslav Andjelković[1,3] & Bosiljka Tadić[1]

Multilevel self-assembly involving small structured groups of nano-particles provides new routes to development of functional materials with a sophisticated architecture. Apart from the inter-particle forces, the geometrical shapes and compatibility of the building blocks are decisive factors. Therefore, a comprehensive understanding of these processes is essential for the design of assemblies of desired properties. Here, we introduce a computational model for cooperative self-assembly with the simultaneous attachment of structured groups of particles, which can be described by simplexes (connected pairs, triangles, tetrahedrons and higher order cliques) to a growing network. The model incorporates geometric rules that provide suitable nesting spaces for the new group and the chemical affinity of the system to accept excess particles. For varying chemical affinity, we grow different classes of assemblies by binding the cliques of distributed sizes. Furthermore, we characterize the emergent structures by metrics of graph theory and algebraic topology of graphs, and 4-point test for the intrinsic hyperbolicity of the networks. Our results show that higher Q-connectedness of the appearing simplicial complexes can arise due to only geometric factors and that it can be efficiently modulated by changing the chemical potential and the polydispersity of the binding simplexes.

Self-assembly of nanoscale objects has been recognised as a powerful method enabling the design of advanced materials with new optical, magnetic, conducting and other properties[1–4]. Complex materials with a new functionality often exhibit hierarchical architecture[5–8], suggesting that the self-assembly occurs at different scales from individual nanoparticles to groups and clusters to macroscale materials. In this regard, a cooperative binding of small formatted nanoparticle structures is crucial for the developing large-scale aggregates. They can be prefabricated nanocrystals, self-replicated information-bearing patterns[8–10], or spontaneously formed groups of nanoparticles[1,3,5,11]. The affinity of nanoparticles to merge into a small formation, which then appears as a building block on a larger scale, depends on the particle density and constraints applied in the manufacturing process, and other factors that influence the interactions between them[6,7,12]. In addition to binding energy, this process is regulated by pertinent geometric rules[13–15]. Therefore, the control of the impact of self-assembly at various levels on the emerging hierarchical structure is essential for the new functionality of macroscopic materials. Here, we use numerical modeling to deepen the understanding of cooperative processes of self-assembly and geometric properties of structures that can arise. In the transition from clusters to the solid state, the clusters with a variable size and shape may appear, depending on the materials in question. For example, various atomic clusters represent the energy-minimum configurations of the electronic structure of interacting atoms. Our study is primarily motivated by the nanoparticle self-assembly where nanoparticles of different chemical composition, dimension and morphology can be manufactured. In this case, the energy favourable equilibrium states result through a combination of different inter-particle forces. For instance, an attractive van der Waals component can be rationally balanced by the repulsive electrostatic interaction, additionally modified by pH conditions of the solvent, which results in a variety of precursor clusters[16]. Instead of considering a particular type of interaction, our objectives are to develop a more general mathematical framework that takes into account the impact of the geometric compatibilities of building blocks on the self-assembly process and provides a language to describe the hierarchical structure of the assemblies. The emphasis is on their hidden geometries, which can offer a hint to understand the synergistic effects of the components.

[1]Department of Theoretical Physics, Jožef Stefan Institute, 1000, Ljubljana, Slovenia. [2]Institute of Physics, University of Belgrade, 11080, Belgrade, Serbia. [3]Institute of Nuclear Sciences Vinča, University of Belgrade, 1100, Belgrade, Serbia. Correspondence and requests for materials should be addressed to B.T. (email: bosiljka.tadic@ijs.si)

In this context, a suitable presentation by graphs or nanonetworks[17] enables the use of advanced graph theory methods to elucidate the structure and abstract essential geometrical descriptors of nano-structured materials[15]. For instance, the network model and topology analysis have proved useful in revealing the structural elements that are responsible for the improved tunneling conduction in self-assembled nanoparticle films[18–20], and to identify the hidden order in amorphous materials[21]. Some recent investigations show how the use of topology can open new ways for designing materials inspired by mathematics[15,22]. On the other hand, the research of growing complex networks has recently been extended to explore the attachment of objects (loops, simplexes) under geometric rules and control parameters[23–26]. In this regard, the self-assembly can be understood as a language that can describe the complex architecture of these networks. Varying the assembly rules and parameters enables us to explore a broad range of structures, compared to the laboratory experiments and the potential limits of the aggregation process, and understand the emergence of new features[24–27]. A particular anisotropy of the interaction and spatial constraints can lead to some interesting low-dimensional assemblies, for example, chains[28] and patterns obtained by tiling or recognition-binding on a two-dimensional lattice[24], and self-assembly of loops under the planar graph rules[23]. By contrast, self-assembly of geometric objects without spatial embedding can lead to complex, hierarchically organized networks. Beyond the standard graph-theory metric[29,30], the advanced techniques of algebraic topology of graphs[31,32] are used to explore the hidden topology of these networks; the primary goal is to find out how different geometric elements (simplexes) are mutually combining to make simplicial complexes. Analysis based on algebraic topology of graphs has been used in some recent studies, for example, to describe the hierarchical organization of social graphs[33] and the structure of the phase-space manifolds near the jamming transition[34], as well as to adequately quantify the patterns of inter-brain coordination[35] and logically structured knowledge networks[36]. Moreover, in the hidden geometry metric of many complex networks, the closeness of the nodes is expressed by the graph's generalization of negative curvature or hyperbolicity[25,26,37]. It plays a significant role in the network's function. For example, a direct survey of the related graphs revealed the impact of negative curvature on metabolic processes[38] and traffic on the Internet[39].

Here we introduce a model for the cooperative self-assembly, in which small, ordered structures of particles can be recognized as *simplexes* or full graphs (cliques) of different size that attach by nesting in a growing network. The process depends on the size of the group that is formed by the attachment, and it is directed by two ingredients. These are *geometric factor*, which refers to the availability of the geometrically appropriate sites where the clique can nest along one of its lower-dimensional faces, and the *chemical factor* associated with the affinity of the system for simultaneously binding an excess number of particles. We notice that for a simplex of a given size the geometric constraints change systematically when the network grows, whereas the chemical affinity affects the actual binding. By exploiting the interplay of these elements, we develop various classes of assemblies represented by graphs, and we investigate their structure using graph-theoretic metrics. We show that these structures possess higher combinatorial connectivity, which can be quantified by algebraic topology measures. With a large number of examples, we demonstrate how the geometrical element that plays a vital role in the appearance of the higher Q-connectedness can be enhanced or reduced by changing the chemical affinity of the assembly. We also show that these new structures exhibit a global negative curvature or δ-hyperbolicity. Our model concerns poly-dispersive cliques, whose size varies according to a given distribution in the range from a connected pair of nodes to 12-clique. As a particular case, we consider the aggregation of mono-disperse cliques of a given order. Below are the details of the model explained in the formal language of topology; to present the model at work, we provide the Web applet[40].

Results and Discussion

Computational model. A clique of order q_{max} is fully connected graph of $s = q_{max} + 1$ nodes; some examples are shown in Fig. 1. Faces of the clique are cliques of the lower orders which are contained in the original clique $\sigma_q \in \sigma_{q_{max}}$, where $q = 0, 1, 2, 3 \cdots q_{max} - 1$ correspond to a single node, a pair of connected nodes, a triangle, tetrahedron, etc., up to the largest subgraph of the order $q_{max} - 1$, respectively. The number of equivalent faces is given by $C_q = \begin{pmatrix} q_{max} + 1 \\ q + 1 \end{pmatrix}$. Hence, $C_0 = q_{max} + 1 = s$ is the dimension or the number of nodes involved in the considered clique. We assume that by docking (or nesting), a clique *shares a face of order q* with an already existing clique in the networks. In this way, the number of simultaneously added particles (nodes) is given by the difference of the dimension of the clique to be formed by docking and the size of the docking site, i.e., $n_a = (q_{max} + 1) - (q + 1)$. Furthermore, we assume that the system's affinity ν towards adding new particle is finite. Therefore, the probability of docking along a particular face of the clique is weighted by the factor $e^{-\nu n_a}$, considering the complementary n_a particles. Therefore, the normalized probability for docking of a clique of order q_{max} along its face of order q is given by

$$p(q_{max}, q; t) = \frac{c_q(t) e^{-\nu(q_{max} - q)}}{\sum_{q=0}^{q_{max}-1} c_q(t) e^{-\nu(q_{max} - q)}}, \tag{1}$$

where $c_q(t)$ is the number of geometrically similar docking sites of the order q at the evolution time t. In our model, a clique is formed in each time step t; the size of the clique can vary in a given range. In particular, here we consider cliques of the dimension $s \in [1,12]$ taken from a power-law distribution $g(s) = As^{-2}$, where A is the corresponding normalisation factor. The empirical fact motivates this form of the distribution, namely, that larger cliques appear less often in modular networks. The network growth by addition of mono-disperse cliques is a particular case of our model. For instance, by fixing $s_{min} = s_{max} = 3$ (triangles) and $s_{min} = s_{max} = 4$ (tetrahedra), we obtain two types of networks with mono-disperse cliques.
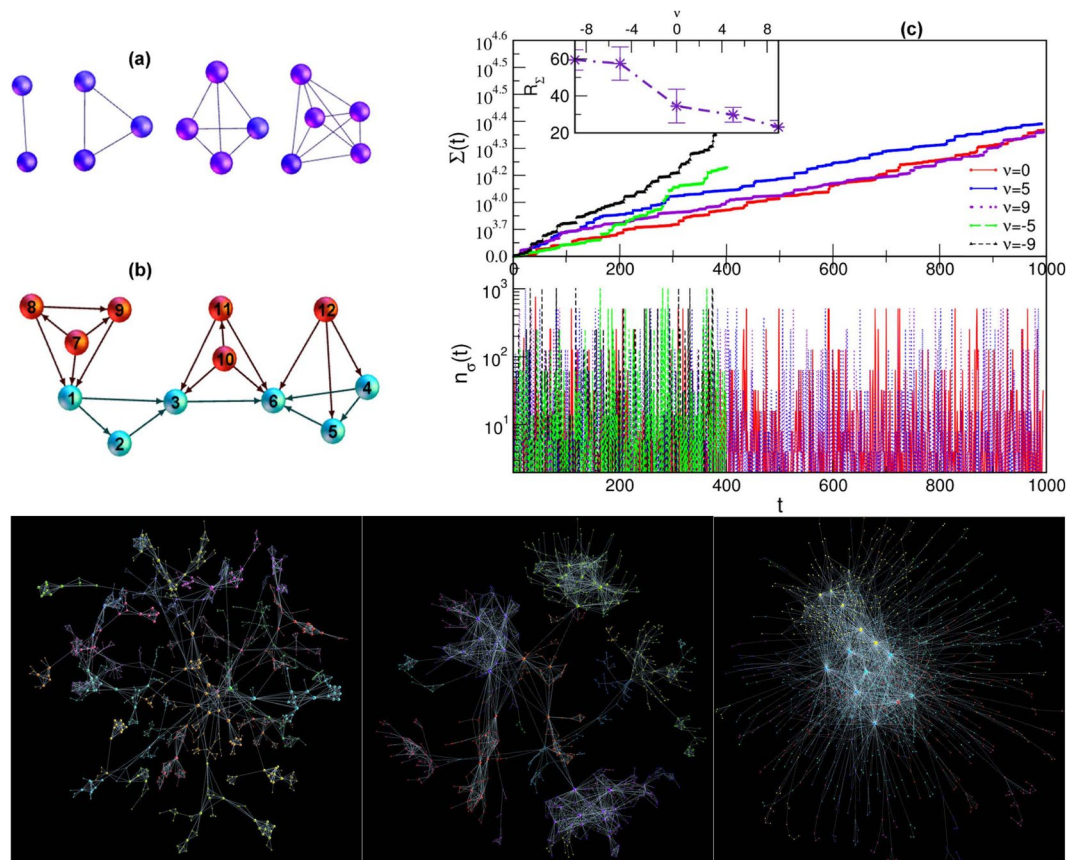
Figure 1. (**a**) Examples of geometrical shapes identified as cliques of the order $q_{max} = 1, 2, 3, 4$, from left to right. (**b**) Addition of a tetrahedron ($q_{max} = 3$) to the system of blue nodes can be nested in three different ways, i.e., by its face of the dimension $q = 0$, for instance, including the node "1", $q = 1$, the link "3–6", and $q = 2$, the triangle "4–5–6". The corresponding number of new particles $n_a = q_{max} + 1 - (q + 1) = q_{max} - q$ are shown by red nodes. (**c**) The number of simplexes $\Sigma(t)$ as function of time for aggregation of poly-disperse cliques at different parameter ν. Lower panel shows the corresponding number $n_\sigma(t)$ of added simplexes per time step. Inset: Average growth rate $R_\Sigma \equiv \langle d\Sigma/dt \rangle$ vs. ν. Bottom panels: Networks of aggregated cliques of sizes $s \in [2,10]$ for the varied chemical potential $\nu = -9$, $\nu = 0$, and $\nu = +9$, left to right. Different colours of nodes indicate the network's community structure.

The first clique taken from the distribution $g(s)$ is assembled and considered as the seed structure. Then, at each step, the size of a new clique $s \in g(s)$ is taken and the clique is formed by attaching the number $s - q - 1$ of new nodes with the selected $q + 1$ nodes on the existing structure. Then *the docking condition requires that these $q + 1$ nodes match a q-face of the new clique*. According to Eq. (1), the selection of the simplex of the order q on the current structure depends on the number of geometrically suitable locations and the corresponding weighting factors. Figure 1b illustrates the effects of the geometrical factor in the example of forming a tetrahedron by attachment of n_a red nodes to the small structure shown by the blue nodes. Considering Eq. (1), the case $\nu = 0$ describes the probability of attachment by geometrical factor alone. In this case, the population of docking sites of the order q determines the likelihood that a new clique will attach by its q-face. On the other hand, the number of docking sites of a given size depends on the actual structure of the network. Note that, by adding a particular clique of the order q_{max} to the system, all its *unshared* faces also appear as new cliques of lower orders. Thus, the number of simplexes fluctuates in time depending not only on the dimension of the clique which is formed in the docking event but also on the size of the actual docking site. It should be noted that while the simplicial complexes grow through the attachment of new cliques via shared faces, the process can not generate holes and cliques of the order larger than the cut-off size s_{max} of the original distribution $g(s)$. In the simulations, we keep track of details constituting each event. For example, the small segment of the output file shown below indicates the time step, current network size, the number of simplexes, order of the added clique, the number of new nodes, and list of all nodes which belong to that clique.
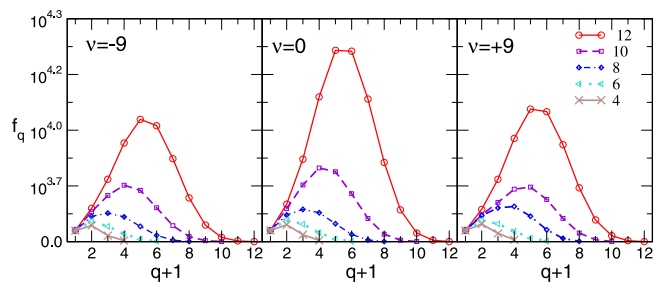
Figure 2. Topological response function f_q plotted against simplex dimension $q+1$ for $\nu = -9$, 0, and $+9$; different curves correspond to the varied upper dimension of the building cliques s_{max} indicated in the legend.

|												
24	42	729	2	1	14	42						
25	45	785	6	3	28	29	32	43	44	45		
26	46	789	3	1	16	28	46					
27	47	791	2	1	12	47						
28	48	795	3	1	17	36	48					
29	49	797	2	1	36	49						
30	55	1805	10	6	17	28	30	31	50	51	52	53	54	55

For varied chemical potential ν, despite the statistically similar population of cliques appearing in the process (taken from the same distribution), the network growth speed and the average rate of the addition of simplexes $R_\Sigma \equiv \langle d\Sigma(t)/dt \rangle$ are different being dependent on the docking probability. Figure 1c displays the evolution of the total number of simplexes $\Sigma(t)$ and the number $n_\sigma(t)$ of the added simplexes per step for different networks until they exceed the target size $N = 1000$ nodes for the first time.

Specifically, a fast growth of the network is observed for the negative values of the parameter ν while much slower growth rates characterize the assembly process for $\nu \geq 0$. In fact, for $\nu < 0$ the system "likes" addition of new particles, which represent the non-nested parts of the new clique. Hence, the cliques effectively repel each other resulting in a sparse structure and fast growth of the network size and also the addition of new simplexes. In contrast, when $\nu > 0$ the cliques are preferably nested along their larger faces, thus reducing the number of the newly added nodes. This powerful attraction among cliques leads to dense network structure and a small number of added nodes and unshared faces per time step. This situation results in a slower growth of the network and reduced simplex addition rate, as shown in Fig. 1c. In contrast, the case with strictly geometrical assembly, $\nu = 0$, has no preference for any size of a docking site; the probability is strictly determined by the number of locations of a given size. Accordingly, these details of the process have an impact onto the topology of the evolving assemblies, which we study in the following. For illustration, three examples of the networks containing the number $N \geq 1000$ nodes for varied parameter ν and the same distribution of the incident cliques are shown in bottom panels in Fig. 1.

Combinatorial topology of aggregates with poly-disperse cliques. As the network examples in Fig. 1 (bottom) demonstrate, the structure that emerges in the assembly of cliques depends strongly on the affinity for the simultaneous attachment of many nodes, apart from the geometrical constraints. Specifically, for large negative values of the parameter ν, an active 'repulsion' between the cliques results in the sparse structure, nearly representing a tree of cliques of different orders. This kind of structures possesses a significant average distance, the modularity, and clustering coefficient, which can be related to the original population of cliques. On the other hand, for the positive values of ν, the cliques firmly attach to each other, resulting in a gradually smaller number of the simultaneously added particles. The appearing structure possesses a large core of densely packed higher-order cliques while low-order structures remain at the periphery. An impressive network architecture with well-separated communities appears for $\nu = 0$, assembled under geometrical constraints alone. As described below, the graph properties are tunned between these extremes by varying the parameter ν.

Here, our focus is on the appearance of higher combinatorial topologies of these graphs, which is directly related to the ways that the assembled cliques share their faces of different orders. In the simulations, we keep track of each added simplex and nodes that participate in it, as explained above. In this way, for a clique of the order q_{max} we can distinguish the number of its shared faces of order $q < q_{max}$. Intuitively, when the groups repel each other, i.e., for $\nu < 0$, their common faces will be the lowest orders, such as single nodes and links and, less often, triangles or higher structures. The opposite situation typically occurs for $\nu > 0$ where the simplexes have a high affinity towards sharing nodes; cf. structure in Fig. 1. Due to shared faces, for instance, of the order q, the number of distinguishable simplexes of that order is smaller than the number of faces C_q of a free added clique. Therefore, the *topological response* function f_q of the network[34] can be determined as the number of different simplexes *at* the topology level q; it provides a good measure of the combinatorial complexity of the assembly in response to the varying external parameters ν, s_{max}. In Fig. 2, we show how the function f_q varies along the topology levels q depending on the parameter ν and the range of the distribution of the attaching cliques. The peak of the distribution shifts towards higher values when larger dimension cliques appear, whereas the height depends on the way that they interconnect at each the topology level. Further, Q-analysis based on the algebraic topology of graphs[32,41–43] is here applied for characterization of the graph architecture by determination of q-connected
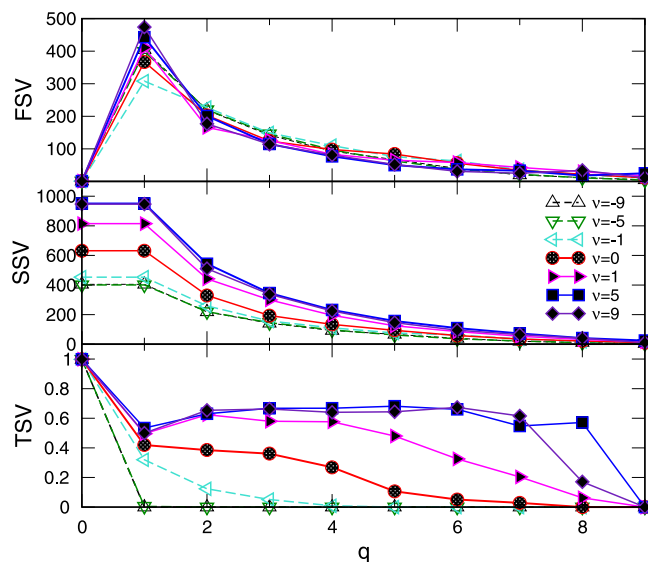
Figure 3. Components of the first (FSV), second (SSV) and third (TSV) structure vector corresponding to the topology level q against q of the networks grown at different values of ν and the fixed distribution of clique size $s \in [2, 10]$.

| q	$\nu = -1$			$\nu = 0$			$\nu = +1$		
	Q_q	n_q	$\hat{Q}_q$	Q_q	n_q	$\hat{Q}_q$	Q_q	n_q	$\hat{Q}_q$
0	1	453	0.997	1	632	0.998	1	815	0.999
1	308	453	0.320	367	632	0.419	411	815	0.495
2	227	259	0.124	203	330	0.385	166	442	0.624
3	149	157	0.051	124	194	0.361	126	300	0.580
4	110	111	0.009	98	134	0.269	83	196	0.576
5	76	76	0.000	84	94	0.106	65	125	0.480
6	63	63	0.000	56	59	0.051	58	86	0.325
7	36	36	0.000	34	35	0.029	43	54	0.204
8	20	20	0.000	22	22	0.000	30	32	0.063
9	11	11	0.000	11	11	0.000	17	17	0.000

Table 1. The components of the three structure vectors for the networks generated at different chemical potential ν.

components for each topology level q. Specifically, for the topology levels $q = 0, 1, 2, \cdots q_{max}$ of each studied network we determine the components of three structure vectors, $\{Q_q\}$, $\{n_q\}$ and $\hat{Q}_q$, defined in Methods. These vectors allow a direct comparison of the hierarchical structure of various emergent networks. In Fig. 3, we plot the components of these structure vectors as a function of q for several assemblies of poly-disperse cliques with different chemical potential ν.

The similarity in the number of q-connected components (FSV) reflects the statistically similar population of cliques of all dimensions (taken from the same distribution) in all studied networks. However, their inherent structure is significantly different, which is expressed by the components of SSV and TSV for various q (see Methods). Notably, the third structure vector in networks for $\nu < 0$ has non-zero components only at lowest topology levels; this implies that different higher-order cliques present in the graph will be separated from each other by removing the structures of the order $q = 1$ (link) between them. The situation is much different in the assemblies grown when $\nu > 0$ where the simplicial complexes containing the higher-order cliques remain strongly interconnected until the before-last level $q_{max} - 1 = 8$. These findings agree with the impact of the chemical potential favoring the cliques attraction for $\nu > 0$ and repulsion for $\nu < 0$. In this context, it is interesting to note that structure that was grown solely under the geometrical rules ($\nu = 0$) already possesses a sizable hierarchical organization of simplicial complexes; although the degree of connectivity is systematically lower than in the case $\nu = +1$, the structure holds together until the level $q = 7$. (See Table 1 for the exact values). As the Fig. 3 shows, this hierarchical architecture of the assembled networks gradually builds with increasing values of the parameter ν. To illustrate the differences in the hierarchical organization of the systems for $\nu = -1, 0, +1$, in Fig. 4 we display those parts of their structure that are still visible at the topology level $q = 5$. Precisely, the nodes participating in the simplexes of order $q \leq 5$ which are not faces of the cliques of the order $q > 5$, are removed. The connections among the remaining nodes are shown according to the network's adjacency matrix.
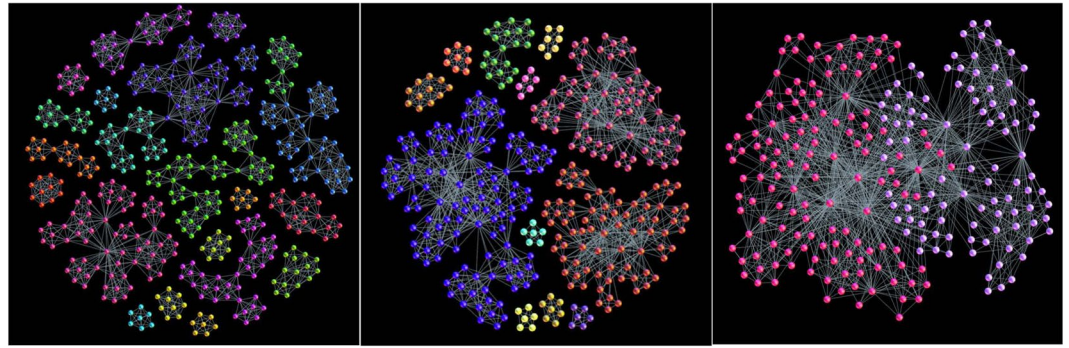
Figure 4. Adjacency matrix of the network's nodes which participate in structures that are "visible" at the topology level q = 5. L Left to right: $\nu = -1$, 0, and +1. Different colors identify clusters or communities.
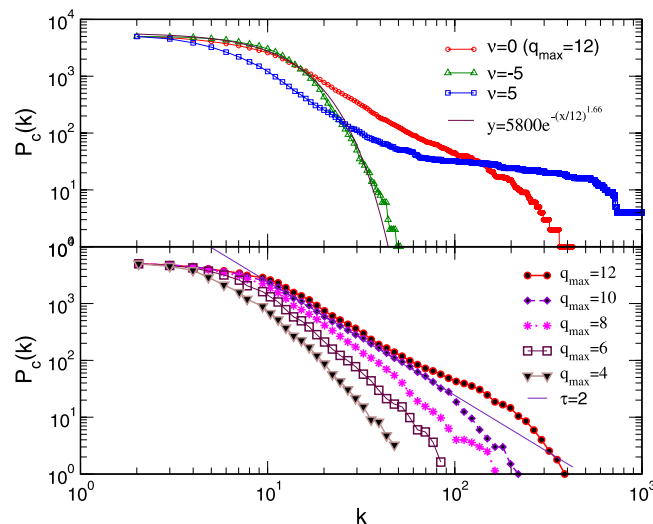

Figure 5. Cumulative distributions of the degree in networks of aggregated poly-disperse cliques $s \in [2, 12]$ and varied chemical potential ν (top panel) and for purely geometrical aggregation ($\nu = 0$) and varied size of the largest added clique s_{max} (lower panel). Each distribution is averaged over several samples of the networks with the number of nodes $N \geq 5000$.

The node's participation in building various simplexes also manifests in the global statistical features of the network. The cumulative degree distribution for several studied aggregates is given in Fig. 5. It is averaged over several realizations of the systems containing over 5000 nodes, where $s_{max} \in [2,12]$. Although a broad distribution of the node's degree occurs in each case, it strongly varies with the parameter ν. It is interesting to note that, in the networks grown by geometrical constraints with $\nu = 0$, we obtain the distribution with a power-law decay $\tau + 1 \approx 3$ (within the numerical error bars); its cut-off appears to depend on the size of the largest clique. In contrast, the exponential decay is observed for $\nu < 0$ while a structure containing many nodes of a large degree is present in the case of clique attraction for $\nu > 0$, which is separated from the low-degree nodes. Other graph theoretic measures also vary accordingly.

δ-Hyperbolicity of the emergent networks. For network structures, δ-hyperbolicity is a generalization of negative curvature in the large[37]. Here, we consider the aggregates of cliques, which are known 0-hyperbolic graphs; therefore, these structures are expected to exhibit this intrinsic property at a larger scale. Following the procedure described in[37], we investigate the 4-point Gromov hyperbolicity of different emergent networks. Specifically, we determine the average hyperbolicity $\langle \delta \rangle$ in comparison to the graph's diameter for $\nu = -5, -1, 0,$ $+1$, and $+5$, by a sampling of 10^9 sets of four nodes, as described in Methods. Considering three different realizations of the network for each ν, we find numerically that δ can take the values $\{0,1/2,1\}$; hence, the maximum value $\delta_{max} = 1$ suggests that these assemblies are 1-hyperbolic. In Fig. 6 (bottom panel) we plot the average hyperbolicity $\langle \delta \rangle$ against the minimal distance d_{min} of the involved pair in the smallest sum $\mathcal{J}$, see Methods. Notably, for all network types $\langle \delta \rangle$ remains bounded at small values. In particular, we find that $\langle \delta \rangle = 0$ for the tree graph of cliques corresponding to $\nu = -5$. Whereas, the hyperbolicity parameter is close to zero in the sparse network of cliques for $\nu = -1$, and slightly increases in the more compact structures corresponding to $\nu = 0$ and $\nu > 0$. Note
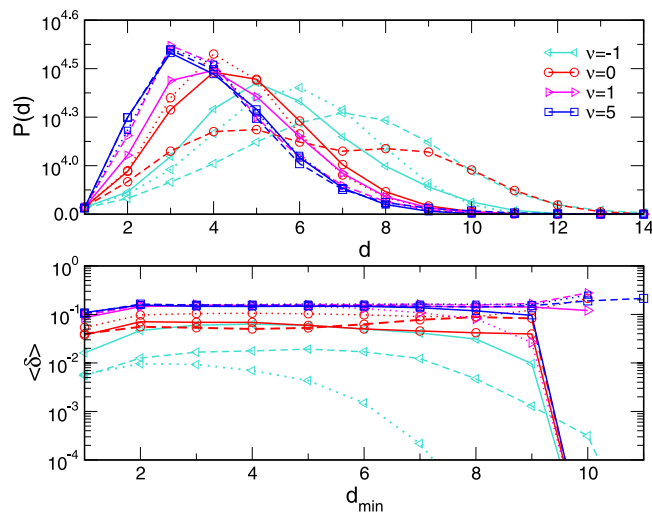
Figure 6. Histograms of shortest distances d between pairs of nodes (top) and average hyperbolicity $\langle\delta\rangle$ vs. d_{min} (bottom) in three samples of networks for $\nu = -1, 0, +1$, and $+5$. Network size is above $N = 500$ nodes.
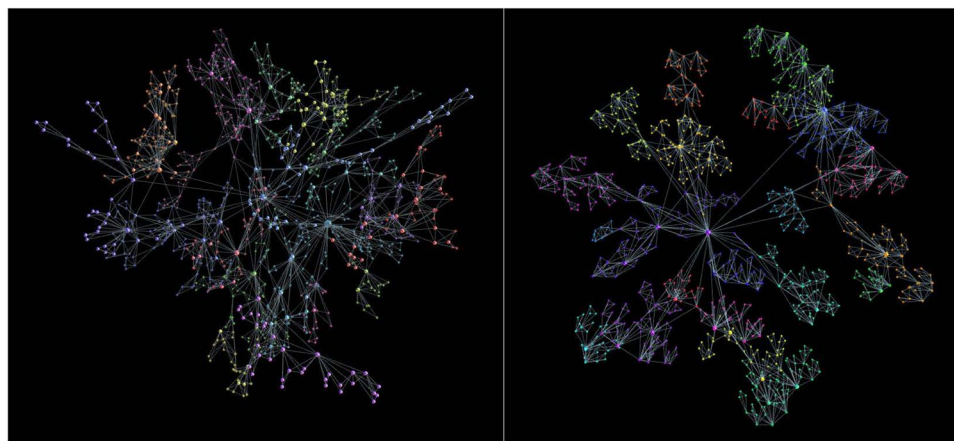
**Figure 7.** Aggregates of tetrahedra (left) and triangles (right) for $\nu = 0.0$.

that due to a small number of pairs of nodes having the largest distance in the graph we observe the fluctuation of $\langle\delta\rangle \in [0, 0.5]$. The histograms of distances between all pairs of nodes in the considered networks are also shown in Fig. 6 (top panel).

Aggregation of monodisperse cliques. In this section, we briefly consider the structures grown with the same aggregation rules but with mono-disperse building blocks. Some compelling examples are the aggregates of tetrahedra and triangles. Tetrahedral forms are ubiquitous minimum-energy clusters of covalently bonded materials[12]. We also study the impact of the chemical potential in the event of aggregation of triangles. The importance of triangular geometry was recently pointed in the context of quantum networks[44]. Some examples of these structures grown by the aggregation rules of our model are shown in Fig. 7.

Since the aggregation process does not alter the size of the largest clique, these networks have only few topology levels. Specifically, in the aggregates of tetrahedra $q_{max} = 3$, and they can share nodes, links, and triangles as faces of lower orders; for triangles, $q_{max} = 2$ and shared faces are links and nodes. Therefore, their structure vectors are rather short. However, they possess a captivating structure of simplicial complexes, depending on the chemical potential and geometry constraints. Consequently, the degree distributions are altered by changing ν, as shown in Fig. 8. Notably, the appearance of some scale-invariant structures is favored by the mutual attraction of cliques for $\nu > 0$. The aggregation of tetrahedra more efficiently builds such structures as compared with triangles. Whereas, the scale-free range is limited with the exponential cut-offs in the case of triangles unless ν is sufficiently large. Further analysis of these and other networks of mono-disperse simplexes is left for future work.

Discussion

We have introduced a computational model for cooperative self-assembly where small, formed groups of particles appear as building blocks for a large-scale structure. In this context, in addition to the binding forces, the geometric constraints exerted by the rising architecture play an important role on the proper nesting of the added block.
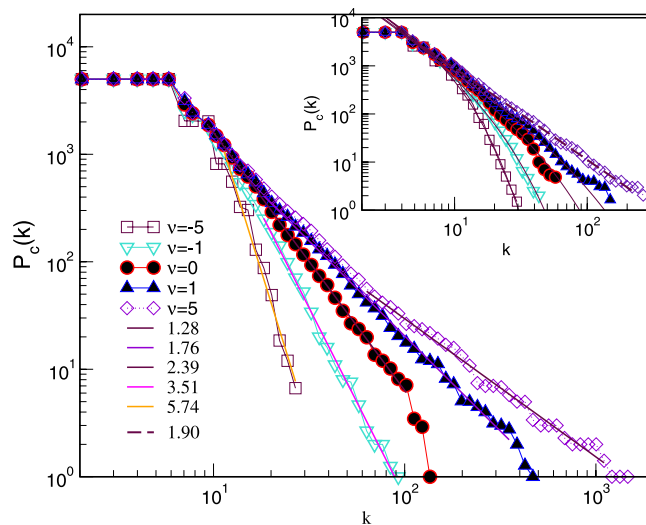
Figure 8. Cumulative distributions of the node's degree in networks of aggregated mono-disperse cliques (main panel) tetrahedra, and (inset) triangles, for different values of the chemical potential ν. Sample averaging and the number of nodes $N \geq 5000$ applies. Thick broken and full lines indicate the range where the slopes given in the legend are measured (within the maximal error bars ± 0.07).

Different geometrically suitable options for nesting a given block structure are further altered due to the chemical affinity ν of the system for receiving the excess number of particles. Formal rules of the model are motivated by situations that usually occur in self-assembly of nanoparticles, where the possibilities for creating different clusters are tremendous. Nevertheless, the rules can be easily adapted to describe different other cases where, for example, due to interactions, only clusters of a certain type can appear and then combine into a hierarchical network.

It should be noted that the model explicitly does not take into account the effects of temperature and diffusion, which are experimentally controllable parameters. In the assembly, simplexes are added one by one, and every added object is attached (with probability one) to the structure when a geometrically suitable nesting site is found and remains in place. Therefore, in the limits described below, these aggregation characteristics resemble the well-known processes of *diffusion limited aggregation* (DLA), where the random particle tree grows in low-density conditions by attaching a particle that diffuses in the solvent when it approaches the tree[45]. Indeed, within the limits of the significant negative values of the ν parameter that promotes repulsive interactions between simplexes, the structure resembles a DLA tree, but here it is made from expanded objects (simplexes) and not individual particles. Note that in this case, ν refers to the number of *excess* particles of the coming simplex, while simplex joins the structure along the nest, containing the remaining particles. Hence, effectively, the chemical potential for the nested particles of the simplex is positive, in analogy to DLA binding. More specifically, when $q_{max} = 1$, only one particle can be added with its link, and the growing structure is a random tree, independent of the ν value (see the web demo[40]). In this case, only one type of binding process occurs with a probability one in the equation (1), regardless of the value and the sign of the parameter ν. For $q_{max} > 1$, however, there are several types of bindings that are differentiated by exponential factors in the formula (1) as described above. Consequently, the emerging structure builds non-random features that are different from DLA clusters, as described in the Results section. We have demonstrated how different assemblies with a complex architecture can be formed in the interplay of these geometric and chemical factors. Moreover, the systematic mapping of the developing structure to the graph not only helps us formally implement the self-assembly process but also provides ways to adequately investigate the new structure employing advanced graph theory and algebraic topology methods.

It is interesting that the complex structure of the assembly that possesses combinatorial topology of higher order can arise due to only geometric factors. These topology features are further enhanced in chemically enforced compaction, and, on the contrary, are gradually reduced in sparse networks resulting from chemically favored repulsion between building blocks. Moreover, depending on the dispersion of the components and the chemical factor, the new assemblies may possess scale invariance and an intrinsic global negative curvature; these features are essential for their practical use and functionality. Our model with graph-based representation provides a better insight into the mechanisms that drive the assembly of hierarchically organized networks with higher topological complexity, which is a growing demand for technological applications.

Methods

Program flow for clique aggregation. *Q-analysis: definition of structure vectors.* To describe the global graph's connectivity[40] at different topology levels $q = 0, 1, 2 \cdots _{max}$, Q-analysis uses notation from algebraic topology of graphs[32,41–43]. Specifically, the first structure vector Q_q represents the number of q-connected components and the second structure vector n_q is defined as the number of simplexes of the order greater than or equal to q. In this context, two simplexes are q-connected if they share a face of the order q, i.e., they have at least $q + 1$ shared nodes. Then the third structure vector determined as $\hat{Q}_q \equiv 1 - Q_q/n_q$ measures the degree of connectivity at the topology level q among the higher-order simplexes. From the adjacency matrix of a considered graph, we

construct incidence matrix by Bron-Kerbosch algorithm[46], where simplexes are identified as maximal complete subgraphs (cliques). Then the dimension of the considered simplicial complex equals the dimension of the largest clique $q_{max} + 1$ belonging to that complex.

Measure of curvature: δ-hyperbolicity definition. Following the studies in[37] and references there, we implement an algorithm which uses the Gromov's hyperbolicity criterion. Specifically, for an arbitrary set of four nodes A, B, C, and D, the distances (shortest path lengths) between distinct pairs of these nodes are combined in three ways and ordered. For instance,

$$d(A, B) + d(C, D) \leq d(A, C) + d(B, D) \leq d(A, D) + d(B, C).$$

Algorithm 1. Program Flow: Growth of the graph by attaching simplexes.

1: **INPUT:** ν, s_{min}, s_{max}, N_{max}
2: initialise graph $\mathcal{G}$ as a simplex of size s taken from the distribution $p(s) = As^{-2}$, which is defined in the range $[s_{min}, s_{max}]$
3: **while** $N < N_{max}$ **do**
4:      select new simplex size $s_{new} \in [s_{min}, s_{max}]$ from the distribution $p(s)$
5:      **for all** simplexes $\sigma \in \mathcal{G}$ whose order $q_\sigma < s_{new} - 1$ **do**
6:          compute $p_\sigma(s_{new}, q_\sigma) = \exp\{-\nu(s_{new} - q_\sigma - 1)\}$
7:      **end for**
8:      normalise the probability such that $\sum_\sigma p_\sigma(s_{new}, q_\sigma) = 1$
9:      select the docking site $\sigma \in p_\sigma(s_{new}, q_\sigma)$
10:     form a new simplex σ_{new} by attaching $s_{new} - q_\sigma - 1$ new nodes to the $q_\sigma + 1$ nodes of the docking simplex σ
11:     sampling the data of interest;
12: **end while**
13: **END**

We denote the largest value $\mathcal{L} = d(A, D) + d(B, C)$, the middle $\mathcal{M} = d(A, C) + d(B, D)$, smallest $\mathcal{J} = d(A, B) + d(C, D)$, and the smallest pair distance of $\mathcal{J}$ as $d_{min} = min\{d(A, B), d(C, D)\}$. Then the graph is δ-hyperbolic if there is a fixed value δ for which any four nodes of the graph satisfy the 4-point condition:

$$\frac{\mathcal{L} - \mathcal{M}}{2} \leq \delta. \tag{2}$$

There is a trivial upper bound $(\mathcal{L} - \mathcal{M})/2 \leq d_{min}$. Hence, by plotting $(\mathcal{L} - \mathcal{M})/2$ against d_{min} we can investigate the worst case growth of the function. For a given graph, we first compute the matrix of distances between all pairs of nodes; then, by sampling a large number of sets of nodes for the 4-point condition (2) we determine and plot the average $\langle \delta \rangle$ against the corresponding distance d_{min}.

Graphs visualisation We used gephi.org for graph presentation and community structure detection by maximum modularity method[47].

References

1. Pelaz, B. *et al*. The state of nanoparticle-based nanoscience and biotechnology: Progress, promises, and challenges. *ACS Nano* **6**, 8468–8483 (2012).
2. Whitesides, G. M. & Grzybowski, B. Self-assembly at all scales. *Science* **295**, 2418–2421 (2002).
3. Boles, M. A., Engel, M. & Talapin, D. V. Self-assembly of colloidal nanocrystals: From intricate structures to functional materials. *Chemical Reviews* **116**, 11220–11289 (2016).
4. Meledandri, C. J., Stolarczyk, J. K. & Brougham, D. F. Hierarchical gold-decorated magnetic nanoparticle clusters with controlled size. *ACS Nano* **5**, 1747–1755 (2011).
5. Wang, L., Xu, L., Kuang, H., Xu, C. & Kotov, N. A. Dynamic nanoparticle assemblies. *Accounts of Chemical Research* **45**, 1916–1926 (2012).
6. Luo, D., Yan, C. & Wang, T. Interparticle forces underlying nanoparticle self-assemblies. *Small* **11**, 5984–6008 (2015).
7. Liu, S. & Yu, J. Cooperative self-construction and enhanced optical absorption of nanoplates-assembled hierarchical Bi2WO6 flowers. *Journal of Solid State Chemistry* **181**, 1048–1055 (2008).
8. Wang, R. *et al*. Self-replication of information-bearing nanoscale patterns. *Nature* **478**, 225–229 (2011).
9. Rossi, L. *et al*. Cubic crystals from cubic colloids. *Soft Matter* **7**, 4139–4142 (2011).
10. Gu, Y. *et al*. Collective alignment of nanorods in thin newtonian films. *Soft Matter* **9**, 8532–8539 (2013).
11. Toulemon, D. *et al*. Enhanced collective magnetic properties induced by the controlled assembly of iron oxide nanoparticles in chains. *Adv. Funct. Mat.* **26**, 1616–3028 (2016).
12. Xing, X. *et al*. Probing the low-energy structures of aluminum-magnesium alloy clusters: a detailed study. *Phys. Chem. Chem. Phys.* **18**, 26177–26183 (2016).
13. Krivovichev, S. V. Combinatorial topology of salts of inorganic oxoacids: zero-, one- and two-dimensional units with corner-sharing between coordination polyhedra. *Crystallography Reviews* **10**, 185–232 (2004).
14. Senyuk, B., Liu, Q., Bililign, E., Nystrom, P. D. & Smalyukh, I. I. Geometry-guided colloidal interactions and self-tiling of elastic dipoles formed by truncated pyramid particles in liquid crystals. *Phys. Rev. E* **91**, 040501 (2015).
15. Kotani, M. & Ikeda, S. Materials inspired by mathematics. *Science and Technology of Advanced Materials* **17**, 253–259 (2016).
16. Trefalt, G., Tadić, B. & Kosec, M. Formation of colloidal assemblies in suspensions for Pb(Mg1/3Nb2/3)O3 synthesis: Monte Carlo simulation study. *Soft Matter* **7**, 5566–5577 (2011).
17. Živković, J. & Tadić, B. Nanonetworks: The graph theory framework for modeling nanoscale systems. *Nanoscale Systems MMTA* **2**, 30–48 (2013).
18. Blunt, M. O. *et al*. Charge transport in cellular nanoparticle networks: Meandering through nanoscale mazes. *Nano Letters* **7**, 855–860 (2007).

19. Šuvakov, M. & Tadić, B. Modeling collective charge transport in nanoparticle assemblies. *Journal of Physics: Condensed Matter* **22**, 163201 (2010).
20. Tadić, B., Andjelković, M. & Šuvakov, M. The influence of architecture of nanoparticle networks on collective charge transport revealed by the fractal time series and topology of phase space manifolds. *Journal of Coupled Systems and Multiscale Dynamics* **4**, 30–42 (2016).
21. Hirata, A. *et al*. Geometric frustration of icosahedron in metallic glasses. *Science* **341**, 376–379 (2013).
22. Ikeda, S. & Kotani, M. *A new direction in mathematics for materials science. Springer Briefs in the Mathematics of Materials* **1** (Springer, Tokyo, 2015).
23. Šuvakov, M. & Tadić, B. Topology of Cell-Aggregated Planar Graphs. Computational Science–ICCS 2006: 6th International Conference, Reading, UK, May 28-31, 2006. Proceedings, Part III, Alexandrov, V. N., van Albada, G. D. & Sloot, P. Editors (Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 1098–1105 2006).
24. Tesoro, S., Göpfrich, K., Kartanas, T., Keyser, U. F. & Ahnert, S. E. Nondeterministic self-assembly with asymmetric interactions. *Phys. Rev. E* **94**, 022404 (2016).
25. Bianconi, G. & Rahmede, C. Emergent hyperbolic network geometry. *Sci. Rep.* **7**, 41974 (2017).
26. Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A. & Boguñá, M. Hyperbolic geometry of complex networks. *Phys. Rev. E* **82**, 036106 (2010).
27. Šuvakov, M. & Tadić, B. Transport processes on homogeneous planar graphs with scale-free loops. *Physica A: Statistical Mechanics and its Applications* **372**, 354–361 (2006).
28. Toulemon, D. *et al*. Enhanced collective magnetic properties induced by the controlled assembly of iron oxide nanoparticles in chains. *Advanced Functional Materials* **26**, 2454–2462 (2016).
29. Bollobás, B. *Modern Graph Theory* (Springer, New York, inc, 1998).
30. Dorogovtsev, S. N. Lectures on Complex Networks. *Oxford Master Series in Statistical, Computational, and Theoretical Physics*. (Oxford University Press, Oxford, 2010).
31. Kozlov, D. Combinatorial Algebraic Topology. *Springer Series Algorithms and Computation in Mathematics* 21 (Springer-Verlag, Berlin, Heidelberg, 2008).
32. Jonsson, J. Simplicial Complexes of Graphs. *Lecture Notes in Mathematics*. (Springer-Verlag, Berlin, 2008).
33. Andjelković, M., Tadić, B., Maletić, S. & Rajković, M. Hierarchical sequencing of online social graphs. *Physica A: Statistical Mechanics and its Applications* **436**, 582–595 (2015).
34. Andjelković, M., Gupte, N. & Tadić, B. Hidden geometry of traffic jamming. *Phys. Rev. E* **91**, 052817 (2015).
35. Tadić, B., Andjelković, M., Boshkoska, B. M. & Levnajić, Z. Algebraic topology of multi-brain connectivity networks reveals dissimilarity in functional patterns during spoken communications. *PLoS ONE* **11**, e0166787 (2016).
36. Andjelković, M., Tadić, B., Mitrović Dankulov, M., Rajković, M. & Melnik, R. Topology of innovation spaces in the knowledge networks emerging through questions-and-answers. *PLoS ONE* **11**, e0154655 (2016).
37. Kennedy, W. S., Saniee, I. & Narayan, O. On the hyperbolicity of large-scale networks and its estimation. In *2016 IEEE International Conference on Big Data (Big Data)*, IEEEXplore, pp. 3344–3351 (2016).
38. Albert, R., DasGupta, B. & Mobasheri, N. Topological implications of negative curvature for biological and social networks. *Phys. Rev. E* **89**, 032811 (2014).
39. Narayan, O. & Saniee, I. Large-scale curvature of networks. *Phys. Rev. E* **84**, 066108 (2011).
40. Šuvakov, M., Andjelković, M. & Tadić, B. Applet: Simplex aggregated growing graph Date of access: 03/01/2018. http://suki.ipb.ac.rs/ggraph/ (2017)
41. Atkin, R. H. An algebra for patterns on a complex, II. *International Journal of Man-Machine Studies* **8**, 483–498 (1976).
42. Johnson, J. Some structures and notation of Q-analysis. *Environment and Planning B: Planning and Design* **8**, 73–86 (1981).
43. Beaumont, J. R. & Gatrell, A. C. *An Introduction to Q-Analysis*. (Geo Abstracts, Norwich-Printed by Edmund Nome Press, Norwich, 1982).
44. Bianconi, G., Rahmede, C. & Wu, Z. Complex quantum network geometries: Evolution and phase transitions. *Phys. Rev. E* **92**, 022815 (2015).
45. Ball, R., Nauenberg, M. & Witten, T. A. Diffusion-controlled aggregation in the continuum approximation. *Phys. Rev. A* **29**, 2017–2020 (1984).
46. Bron, C. & Kerbosch, J. Finding all cliques of an undirected graph. *Comm ACM* **16**, 575–577 (1973).
47. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **10**, P10008 (2008).

Acknowledgements

Author Contributions

B.T. and M.S. designed research, M.S. contributed program tools, M.A., B.T., M.S. performed simulations and analysed data, B.T. produced figures and wrote the manuscript, all authors reviewed the manuscript.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.