

SCIENTIFIC REPORTS



OPEN

A nature inspired modularity function for unsupervised learning involving spatially embedded networks

Raj Kishore¹, Ajay K. Gogineni², Zohar Nussinov³ & Kisor K. Sahu¹

The quality of network clustering is often measured in terms of a commonly used metric known as “modularity”. Modularity compares the clusters found in a network to those present in a random graph (a “null model”). Unfortunately, modularity is somewhat ill suited for studying spatially embedded networks, since a random graph contains no basic geometrical notions. Regardless of their distance, the null model assigns a nonzero probability for an edge to appear between any pair of nodes. Here, we propose a variant of modularity that does not rely on the use of a null model. To demonstrate the essentials of our method, we analyze networks generated from granular ensemble. We show that our method performs better than the most commonly used Newman-Girvan (NG) modularity in detecting the best (physically transparent) partitions in those systems. Our measure further properly detects hierarchical structures, whenever these are present.

One of the most successful strategies in understanding a complex scientific/engineering problem involves two essentially crucial steps: deconstruction and reconstruction. In the “deconstruction” stage, one breaks a complex, nearly intractable problem into many elementary parts. These basic constituents may be then more readily studied. During reconstruction, one glues these smaller parts back together, as they appear in the original system. In reconstruction of the original system from its basic building blocks, one places emphasis on attempting to understand the response of the entire system as a whole. In the past, most deconstruction steps used to be performed manually (a step often somewhat facilitated by taking advantage of inherent symmetries of the problem whenever these were present). With the advancement of technology, the complexity and size of routinely investigated systems are constantly increasing. Consequently abstract and scientifically superior mathematical methods, like network partitioning are becoming increasingly popular. These methods play a more prominent role in the deconstruction step. Networks appear naturally and may efficiently represent numerous systems in diverse fields including the social sciences, biology, information technology, neurosciences and economics. The architectures of real physical spatially embedded networks where the nodes are defined by their positions in Euclidean space are highly influenced by geometric constraints. Such physical networks are fundamentally different from networks that are not anchored in Euclidean space^{1–3}.

In the current work, we will examine a particular class of spatially embedded networks, which are obtained from granular materials. In this type of networks, both long distance edges and hubs (high degree nodes) are absent. Also, the total number of edges in such graphs scales linearly with the number of nodes. The absence of hubs renders the topologies relatively feature-less. This more monotonous structure poses a great challenge for general purpose partitioning algorithms. Network analysis of granular materials and their ensembles often facilitates a scientific description of the complex structure of interacting particles, responsible for many bulk properties like mechanical stability, acoustic and heat transmission^{4–7} and paves the path for automated data analytics tools. The first step in this approach is to represent a granular assembly as a network. Herein, each particle is depicted by a node. An edge between any two particles captures their physical contacts or interactions (e.g. force). The second

¹School of Minerals, Metallurgical and Materials Engineering, Indian Institute of Technology Bhubaneswar, Bhubaneswar, 752050, India. ²School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Bhubaneswar, 752050, India. ³Department of Physics, Washington University in Saint Louis, Saint Louis, MO, 63130-4899, USA. Correspondence and requests for materials should be addressed to K.K.S. (email: kisorsahu@iitbbs.ac.in)

step is to detect “community structures” in these networks by optimizing an objective function^{8,9} to extract any underlying mesoscale architecture.

The quality of partitioning a network, also known as clustering, is reflected in its ability to make a distinction between densely connected nodes of “communities” or “modules” that are sparsely linked to other communities. The most prevalent objective function used for this optimization is the so-called “modularity” of a network. A maximization of the modularity function leads to putative partitions of the network into clusters. Amongst many different modularity functions^{10–13}, Newman-Girvan (NG) modularity^{14,15} is most commonly used for real networks^{16–19}. The NG modularity contrasts the number of actual connections in a given network to that of a null model. The null model is obtained by randomly changing the network connections without altering degree of respective nodes. Conceptually, the construction of the null model can be achieved in the following way: (i) all the edges are cut in two and consequently (ii) any two half-edges are re-connected with equal probability. For single layer network, the NG modularity is given as

$$Q_{NG} = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(\sigma_i, \sigma_j) \quad (1)$$

Here, m is total number of edges in the network, σ_i denotes the community that node i belongs to, the numbers k_i and k_j are the degree of i^{th} and j^{th} node respectively (i.e the number of edges that respectively, have node i and j as an endpoint), A_{ij} is the adjacency matrix ($A_{ij} = 1$, if an edge between i^{th} and j^{th} node exist, and $A_{ij} = 0$ otherwise) and $\delta(\sigma_i, \sigma_j)$ is the Kronecker delta whose value will be unity if i^{th} and j^{th} node have the same community membership and zero otherwise.

Since the NG modularity contrasts the real connections in the given network with the randomized connection of the null model as stated above, it is well suited for the networks where connections between any two arbitrary nodes are, at least in principle, not forbidden. However in many physical systems in which the networks are embedded in finite three- or two- dimensional Euclidian space, the edges are distance dependent. In particular, in many such systems, edges between arbitrary distant sites do not appear (since there are essentially neither forces nor correlations between far separated particles or sites). Thus it seems somewhat inappropriate to penalize (i.e render the value of modularity lower) for the absence of edges between two nodes, which are not in close physical proximity. An edge between distant nodes is not physically possible (even though they might belong to same community). Therefore the null model should be restricted from making those connections for spatially embedded networks. Attempts have been made to adapt the modularity for spatially embedded static and temporal networks, though, invariably, some of the most important works^{3,20–23} uses null models. This is appropriate when, at least some knowledge exist for the systems under investigation.

With the rapid growth of modern technology, sensors and devices collect vast raw data. This engenders “the big data” problem in many different arenas. Often the volume of raw unlabeled data far outpaces the increase in labeled data, as in e.g. medical diagnostics. In many cases, even basic information about the field of study might not be available (or is economically unfavorable to harness for the purpose of supervised learning). It is very likely that unsupervised learning is going to play an increasingly vital role in analyzing “big data”. The quintessential example of unlabeled data systems that we study in this article describes granular materials. To the best of our knowledge, no labeled data of meaningful quantity exists in the public domain for these systems. This lack of labeled data occurs despite the fact that granular materials lie at the core of many industries including pharmaceutical, mining, mineral, chemical, fertilizer and agriculture^{24–26}. Moreover, from the principle of objectivity, one can expect that sufficient information for optimal/nearly-optimal partitioning of a network can be sourced intrinsically.

Based on these considerations, here we propose a new physically inspired function, that is extremely simple and incorporates provisions for geometrical constraints for spatially embedded networks and demonstrate its suitability for use in unsupervised learning methods using granular network as a test case. Our model enables the detection and characterization of structures that vary over a broad range of length scales. As such, our model might be useful for many real world applications in which multiple length scales are significant.

Methods

Development of granular network. Granular networks for the present study have been created using two different protocols. The first protocol uses a physics-based Discrete Element Method (DEM) that simulates true granular dynamics. In the second protocol, we synthetically generate granular ensemble where, particles are positioned in a hexagonal lattice. We then use a square stencil to remove particles and therefore produce a patterned disordered region.

DEM simulation to generate granular ensemble. The granular ensemble of the present study is created using DEM^{27–29} simulation of 7428 macroscopic 2D particles of uniform size (radius 0.01 m) by centripetal packing. In this approach, an externally applied centripetal force (magnitude effectively equal to gravitational force), is directed towards the center of the box. DEM simulation protocol adopted for this article is identical to refs^{30,31}, but nevertheless described in Supplementary Information (SI)-1 for the sake of completeness. All the particles move towards the center of the box because of this centripetal force and form a packing. During this packing process, collisions between particles cause reduction in their kinetic energy. Most of the particles in final structure pack in six-fold coordinated hexagonal structure. A very small fraction of the particles have coordination number (number of physically touching neighbors) different from six (Fig. 1a). The network is constructed by placing an edge between any two granules if they are in physical contact. Since most of the particles in the ensemble display six-fold coordination, the resultant network is a composition of mostly 6-regular sub-graph regions separated by

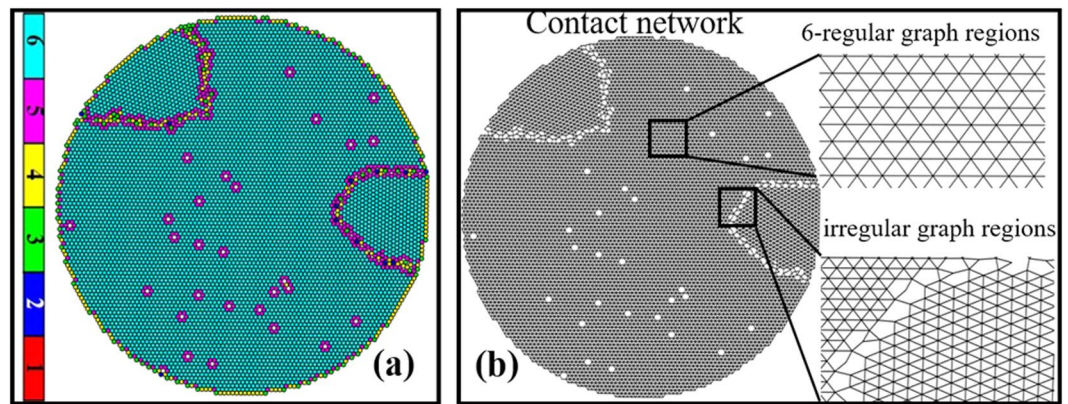


Figure 1. Granular ensemble and corresponding granular network. **(a)** Packing of 7428 uniform sized disks obtained by centripetal packing through DEM simulation (see main text and SI-1). The particles are colored based on their coordination numbers (see the colorbar at the extreme left). The ordered and disordered regions can be visually discerned. A good modularity function should become maximal for the partition that matches these partitions obtained by visual inspection. **(b)** Contact network formed from ensemble **(a)** by following the recipe: an edge is drawn between two particles if they are in physical contact. The weight of all edges is fixed as unity to make it an un-weighted network. Six-regular graph regions and irregular regions are shown in inset (extreme right). [The files containing node connectivity (graph_srep.gml) and the coordinates of each node and their corresponding degree (Particle_pos_with_coord_no_Srep.xls) can be downloaded from <https://sites.google.com/a/iitbbs.ac.in/kks-research-work/research-data>].

irregular regions (Fig. 1b). The irregular regions consist of weakly connected nodes. These regions may be visually discriminated from the strongly connected regular regions.

Synthetically generated granular ensemble. In this protocol, particles are positioned in a regular hexagonal lattice. We then use a square stencil to remove particles and therefore produce a patterned disordered region. These patterns form logical boundaries for community partitions. The network formation protocol from these ensembles is same as discussed earlier.

Finding the communities. For finding and creating communities in the network, we employ two different protocols. In the first protocol, we use a Potts model based community detection scheme discussed next. In the second protocol, we manually set the communities.

Potts model based community detection algorithm. The accuracy of a graph clustering algorithm is mainly controlled by the quality function it uses. It should favor more edges inside community as well as restricts large number of missing edges by using some penalization function. We have used spin-glass-type Potts model algorithm for the community detection of Rhonhovde and Nussinov⁹ (hereafter referred as RN model) This algorithm is described in details in SI-2 and the complete c++ code of RN method can be downloaded from <https://sites.google.com/a/iitbbs.ac.in/kks-research-work/research-data>. The RN method attempts to iteratively find a partition that corresponds to the ground state of the following energy function (or Hamiltonian H) given in Eq. (2).

$$H(\{\sigma\}) = -\frac{1}{2} \sum_{i \neq j} (a_{ij} A_{ij} - \gamma b_{ij} I_{ij}) \delta(\sigma_i, \sigma_j) \quad (2)$$

The asymmetry between connected ($A_{ij} = 1$) and disconnected ($J_{ij} = 1 - A_{ij} = 1$) edges may be reflected by individually setting the respective edge weights (denoted by a and b). The multiplicative factor, γ , is used as a structural resolution parameter. Altering the value of γ may increase or decrease the missing edge weights thus providing control over size and number of communities found by minimizing the Hamiltonian of Eq. (2). Typically, larger values of γ favor smaller communities, and vice versa.

The Potts Hamiltonian favors an intra-community link whereas it disfavors the missing edges within same community. The optimised state of the potts model Hamiltonian have mostly similar spin entities in same state and vice versa³².

Manually setting the communities. In this protocol, we manually set the community boundaries without any regard to its optimality. If the community boundaries coincide with those associated with naturally occurring identifiable structures, clearly representing an optimal solution, then one may expect a higher modularity value, provided the function is properly constructed. At other times, we forcefully set the community boundaries not to coincide with such structures. Then one may anticipate a very low modularity values for such sub-optimal partitions. We will clearly demonstrate that our new modularity function is more sensitive to such changes and outperforms the NG modularity.

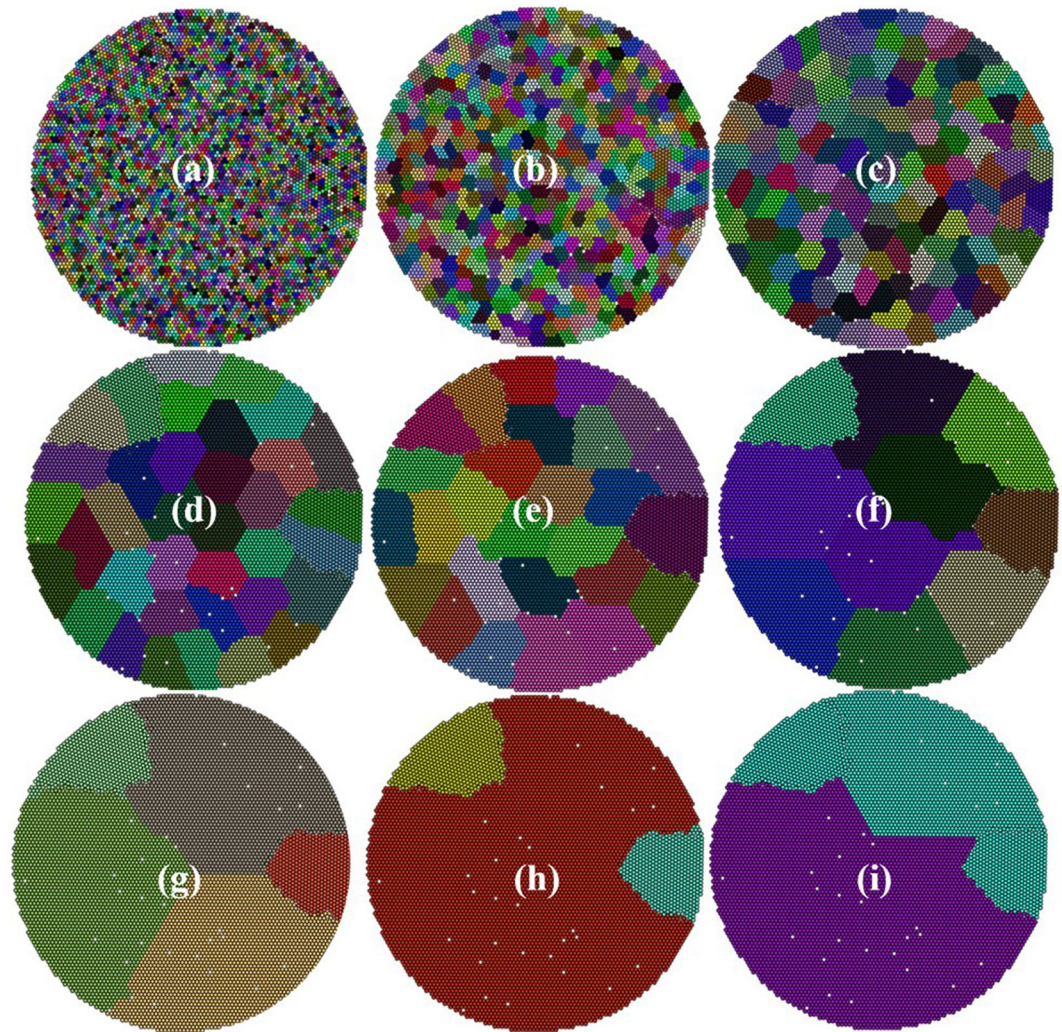


Figure 2. The community structures obtained at different γ values (a) 10 (b) 3×10^{-1} (c) 5×10^{-2} (d) 7×10^{-3} (e) 5×10^{-3} (f) 9×10^{-4} (g) 6×10^{-4} (h) 4×10^{-4} (i) 5×10^{-5} . The edges are not shown for clarity. All the nodes of a community have been assigned same color to make them visually separable and the color for each community is randomly selected. By a visual comparison with Fig. 1a,b, one can immediately point that (h) shows the most optimal partition. In next figure, both NG- and new-modularity functions will be quizzed about the quality of these partitions.

New Modularity function. Before presenting our new function, we briefly discuss physical intuition that underlies our formulation. This intuition is deeply rooted in the study of naturally occurring magnetic materials. Specifically, the modularity function that we will shortly introduce is inspired by the formation of magnetic domains in ferromagnetic materials. In a magnetic material, there are “domains” in which all the moments tend to locally align (or become “polarized”) along the same direction. Such structures are of low energy. We wish to draw an analogy between such magnetic domains and clusters in a graph. Each node in the graph may be viewed as a local magnetic moments or “spin”, (such as that associated with an individual atom). An edge in the graph represents a magnetic interaction between two local moments (atoms/nodes). Along these lines, each community in the graph can indeed be viewed as a magnetic domain. Our modularity function is modeled by a Potts Hamiltonian³³. The construction of this Hamiltonian essentially hinges upon the choices one makes to model the spin-spin interactions for four different scenarios. Adding these interactions together will yield the full Hamiltonian describing our system. The four possible scenarios in this spin dynamic formulation are the following (see SI-3 for details): (i) interaction between spins of identical polarization within same community (ii) interaction associated with a missing edge between nodes within the same community (i.e two nodes of identical spin polarization), (iii) interaction for an edge between two different communities (or two different magnetic domains of different spin polarization) and (iv) interaction representing a missing edge between two nodes from two separate communities. We will use a Heaviside step function to incorporate geometrical constraint associated with the individual items. A complete pictorial guide for all possible scenarios can be found in SI-4. While, the energy of a system (the Hamiltonian) is an extensive quantity, we will make the modularity an intensive parameter by scaling it with the total number of edges. We do so since, if the architecture of the network is relatively

Protocol Combinations (PC)	Key findings	Protocols used for			Objectives
		Ensemble generation	Finding community	Modularity	
PC 1	Best partition	DEM	RN model. γ used.	NG vs. new	Discuss the role of γ in finding the optimal partition. Evaluate NG vs. new modularity for finding the best partition.
PC 2	Best partition	Synthetic (manual)	RN model. γ used.	NG vs. new	Evaluate NG vs. new modularity for finding the best partition.
PC 3	Hierarchical structures	Synthetic (manual)	Manual (forced). γ not used.	NG vs. new	Evaluate NG vs. new modularity for finding hierarchical structures.

Table 1. Usage of different combinations of protocols and organization of their results.

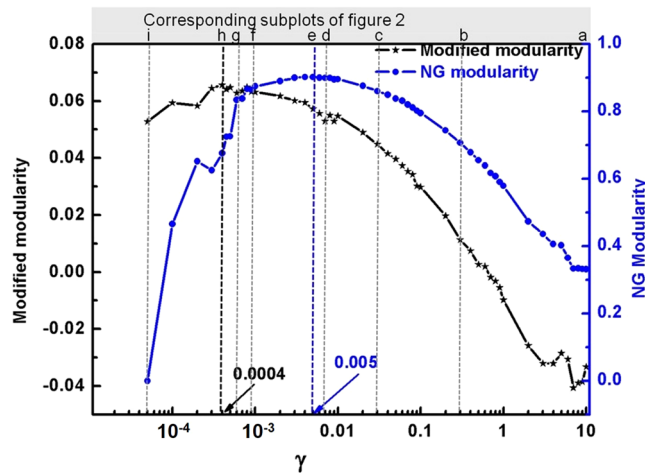


Figure 3. Modularity of community structures obtained at different γ values calculated using NG and our new modularity function of Eq. (3). Our function gives highest value at $\gamma = 0.0004$ (Fig. 2h, the right solution) whereas NG modularity has highest value at $\gamma = 0.005$ (Fig. 2e), clearly not the most optimal solution). The secondary x-axis represents indices of the subplots in Fig. 2.

homogenous, we expect a comparable value of the modularity capturing the system size independent “quality” of the partition. Our proposed modularity function reads:

$$Q(\sigma) = \frac{1}{2m} \sum_{i,j} (a_{ij}A_{ij} - \theta(\Delta x_{ij})|b_{ij}|J_{ij})(2\delta(\sigma_i, \sigma_j) - 1) \tag{3}$$

Here a_{ij} and b_{ij} are the strength (not to be confused with edge weights) of connected and missing edges between i^{th} and j^{th} nodes respectively. For setting up the strengths of edges, a_{ij} and b_{ij} , there can be many choices. We employ a comparison to the local degree distribution between the i^{th} and j^{th} node with the average degree distribution $\langle k \rangle$ of the network, and set

$$a_{ij} = b_{ij} = \left(\frac{k_i + k_j}{2} - \langle k \rangle \right) \tag{4}$$

where, $\langle k \rangle = \frac{1}{N} \sum_{r=1}^N k_r$ and N is total number of nodes and k is the node degree.

$$\{2\delta(\sigma_i, \sigma_j) - 1\} = \begin{cases} 1 & \sigma_i = \sigma_j (i, e \text{ in same community}) \\ -1 & \sigma_i \neq \sigma_j (i, e \text{ in different communities}) \end{cases} \tag{5}$$

Our proposition, therefore, has two important distinctions from almost all of the traditional modularity functions that we know of, including the NG modularity (Eq. (1)) as well as other modularity functions developed for spatial networks. First, it does not explicitly resort to a null model (although the function of Eqs (2 and 3) was largely inspired by such a comparison). Second and most importantly, traditional methods neglect the inter-community interactions (both the existing and missing edges) whereas our function accounts for them. The key difference is embodied by the geometrical dependence that appears in our modularity of Eq. (3). The NG modularity penalizes for all missing edges between two distant nodes within same community. This penalty is imposed even when that edge might not be physically possible because of geometrical constraint. In order to restrict this incorrect over penalization, we introduced, in Eq. (3), a Heaviside unit step function $\theta(\Delta x_{ij})$ which incorporates the geometrical constraints in the form of neighborhood definition. Here $\Delta x_{ij} = x_c - |\vec{r}_i - \vec{r}_j|$ is

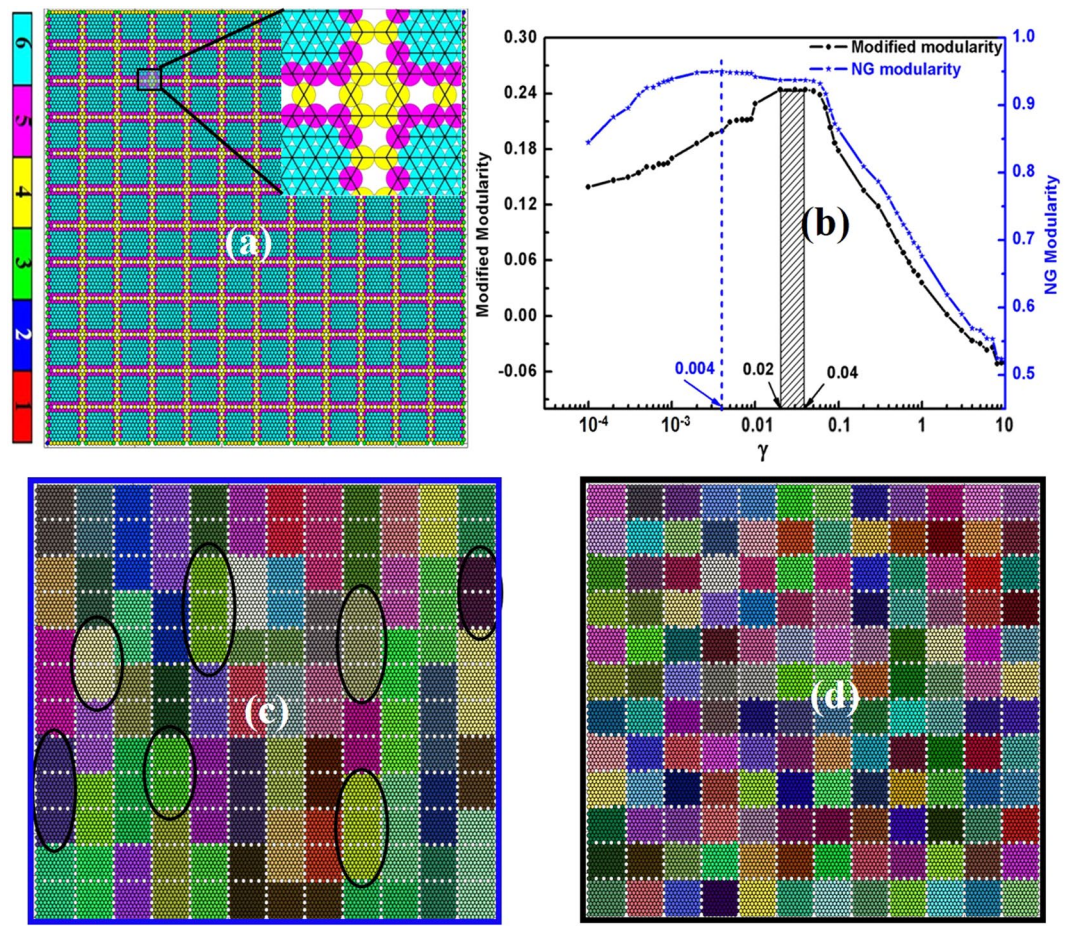


Figure 4. Comparison of best partition obtained using NG modularity vs. new modularity. (a) Synthetically generated granular ensemble with periodic distribution of disordered region. The particles are colored based on their coordination numbers (see the colorbar at the extreme left, which is essentially same as in Fig. 1a). (b) The NG and our new modularity of community structures obtained by RN model at different resolution scale. (c) The maximum modular community structure detected by NG method at $\gamma = 0.004$. It detected many higher-mode ordering as most optimal solution (few of them are encircled; which are clearly not the best solution) (d) Maximum modular community structure detected by our function. This structure is most stable in the range of $0.02 \leq \gamma \leq 0.04$ due to its high modularity. Therefore it is clear that, even in its peak, NG modularity is picking mixed modes (circled regions in c), whereas the new function, in the range of $0.02 \leq \gamma \leq 0.04$, is picking up only the pure primary mode.

the difference in Euclidian distance between nodes i, j and x_c defines cutoff distance for neighborhood and was chosen to be $x_c = 1.05 * (R_i + R_j)$ where, R is the radius of particle for the present study. The function $\theta(\Delta x_{ij})$ introduces a penalty only for those missing links where an edge is geometrically possible.

$$\theta(\Delta x_{ij}) = \begin{cases} 1 & \Delta x_{ij} > 0 \text{ (within cut off)} \\ 0 & \text{otherwise (outside cut off)} \end{cases} \quad (6)$$

Our function compares the local degree distribution at node level with the average degree distribution of the network (Eq. (4)). Its value will be high if the nodes of a community are highly linked with each other. Highly linked communities exhibit a local degree distribution that is larger than the average degree distribution over the entire network. We remark that employing the absolute value $|b|$, in Eq. (3) is not mandatory; this takes care of a subtlety as we explained in SI-4. A thorough discussion on the implications of Eq. (3) for all possible scenarios is also exhaustively discussed therein.

Results

Table 1 outlines our protocol combinations (PCs) and the associated key results. In this study, three types of two-dimensional (2-D) granular networks were prepared using different protocol combinations and their corresponding results are discussed in details. In PC-1 we performed realistic simulation of 2-D granular ensemble (using DEM method) that produced a natural 2-D ensemble (Fig. 1a) from which the network was generated (Fig. 1b). Communities in this network were found using RN method at different resolution values (γ).

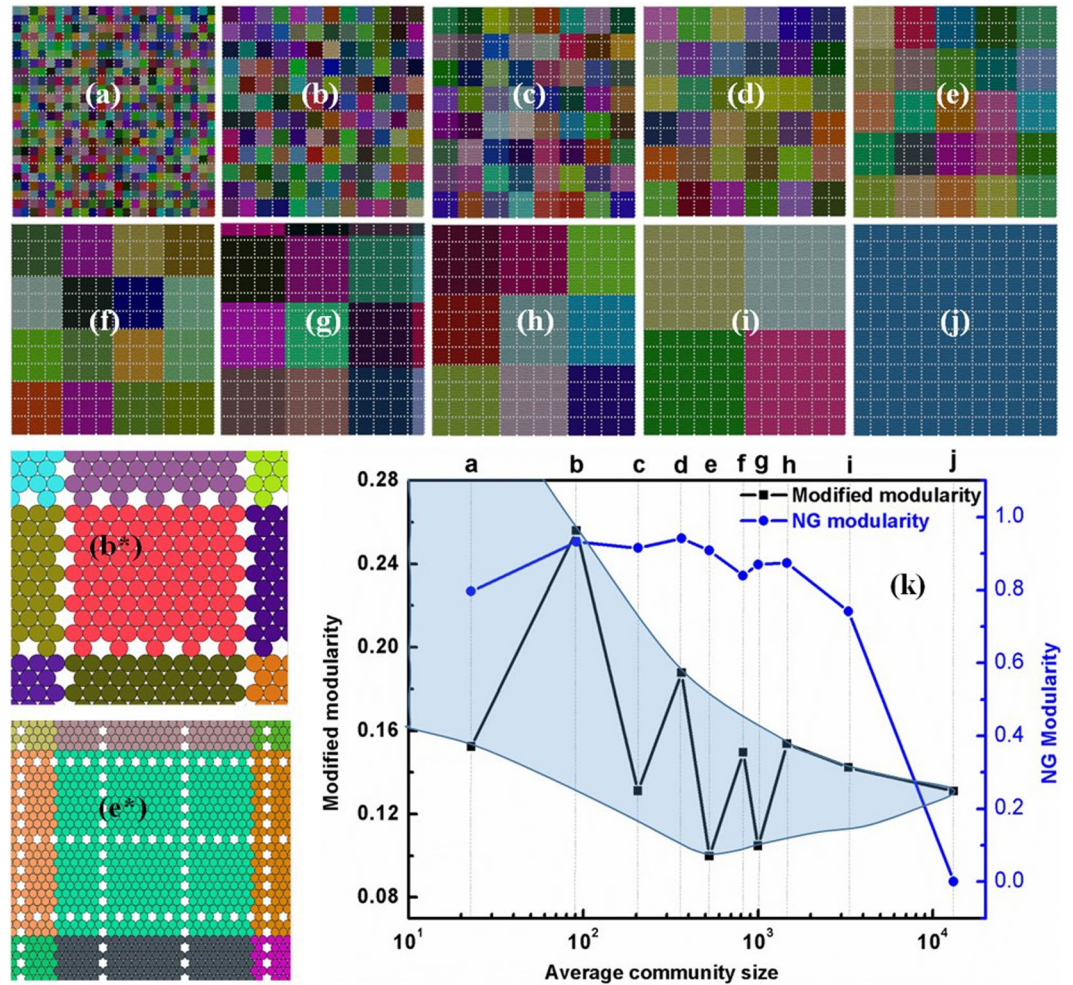


Figure 5. Manually partitioned network (a–j) and corresponding modularity obtained by NG and our new modularity function (as shown in (k)). Our function shows peaks for partition *b*, *d*, *f* and *h* and troughs for *a*, *c*, *e* and *g*; clearly indicating the presence of hierarchical community structures in the given network. Community separation for best (**b**) and worst (**e**) modular structures show that the corresponding partitions can be associated most and least with naturally identifiable structures respectively (*b** and *e** shows a magnified view of portions of *b* and *e* respectively, for clarity). The nature of these two functions is very different (clear from *k*) and demonstrates that, these two functions do not differ by a simple constant.

Modularity for each partition was calculated by both NG- and new- modularity functions. We first illustrate the role of this resolution parameter (γ) and further evaluate the performance of both NG and new modularity functions for finding the best partition. In PC-2, we analyzed a synthetically generated granular ensemble (i.e., without using DEM) with periodic distribution of disordered regions. RN method was used at different resolutions to find the partitions. Both the NG- and new- modularity functions were quizzed about the best partition. The results were compared to “rational human expectations”. In essence, we evaluate, which among these two functions more closely mimic “human-like” vision processing. In PC-3, we again used a synthetically generated granular ensemble. In this case, prime objective is to find which method is more capable in identifying the hierarchical organization. Therefore, the community boundaries were sometime set to coincide with natural boundaries and sometime not. This was done at different levels of hierarchical organizations, i.e., community size was varied from low to high (fine to coarse size).

Results for PC 1. The contact network (Fig. 1b) is created from the granular ensemble (Fig. 1a) as discussed in the ‘DEM simulation to generate granular ensemble’ section of “Method”. The network is then partitioned at different γ values using RN model (Eq. (2)). We have depicted the communities obtained at different γ values (i.e. at varying structural resolution), keeping the node position to be the same as particle coordinates (Fig. 2) to establish a visual correlation between the RN model output and the coordination number plot (Fig. 1a). As the γ reduces in size from 10 to 5×10^{-5} (γ is a dimensionless parameter), the average community size increases gradually. This is reflected in Fig. 2a–i. The γ parameter set the resolution scale. Minimizing Eq. (2) over range of γ values, systematically uncovers organized structures in the network at different length scale. It is vividly seen that Fig. 2h has highest visual correlation with the coordination number plot where the regular/ordered region

are clearly separated from irregular/disordered nodes. Due to this clear demarcation, Fig. 2h is expected to have highest modularity among all the possible partitions. This aspect is investigated in Fig. 3.

It is important to note that the raw values of these two modularity functions (NG vs. our new modified modularity) should not be compared because any one of them can be shifted by an innocuous additive constant. It is the peak location that matters the most. It is evident from Fig. 3 that if community structure is separated at disordered regions then our method finds the most optimal partition (at $\gamma = 0.0004$, corresponding to Fig. 2h), whereas NG method failed to do so (the NG modularity suggested that the best partition occurs at $\gamma = 0.005$; corresponding to Fig. 2e).

Results for PC 2. We next analyzed a synthetically generated granular ensemble with periodic distribution of disordered regions (Fig. 4a; the inset highlighting the network at small scales). This approach creates clearly visible structures, with an in-built hierarchy. The communities were found via the RN model. As seen in Fig. 4b, the NG and our new modified modularity feature similar behavior for high γ values (small average community size). However, for large communities (at lower γ) they behave differently (Fig. 4b).

The best community structure obtained by NG and new modularity is depicted in Fig. 4c,d respectively. The new modularity shows very high visual correlation with natural structure and correctly picks the prime modes only (γ in the range of 0.02–0.04 in Fig. 4b, corresponding partition depicted in Fig. 4d), whereas the NG modularity is picking mixed modes (at $\gamma = 0.004$ Fig. 4b, partition shown in Fig. 4c).

Results for PC 3. Many real world networks contain hierarchical organization where strong groups of vertices have further sub-groups. Granular ensemble in Fig. 5a displays various sub-networks at different length scales. We have manually partitioned this network in roughly equal community sizes, which is so varied as to cover the entire length scale ranging from the primary mode length to system length (Fig. 5a to j, b^* and e^* shows a magnified view of best (b) and worst (e) modular structures). The peaks in our new modularity curve (Fig. 5k) confirm the presence of hierarchical structures in the given network. This illustrates that modular structures at different scales, can be identified by peaks in our modified modularity function. These correspond to the partitions where the community boundaries coincide with disordered regions and therefore help in identifying the higher modes clearly. The most modular structure (Fig. 5b) has all its communities separated through disordered regions (primary mode) whereas the lowest modular (Fig. 5e) has none of its communities separated through the disordered regions. Interestingly, the partition corresponding to Fig. 5f, exhibits a lower NG modularity than both Fig. 5e,g; by contrast, our new modified modularity function shows the opposite but correct trend. It is therefore clear that the NG- and our new- modified modularity functions are characteristically very different (not just by a constant). Our new modularity function is highly sensitive towards coincidence of the disordered regions with the community boundaries as reflected in the large variation in its modularity values (shaded region in modularity curve (Fig. 5k)) whereas NG modularity shows little variation in the range of prime importance. Therefore we establish that, our new modularity function, which is simple in structure and nature inspired, is very capable in finding the best partition, at least for the restricted classes of spherical granular media that are discussed in this article. Further studies are necessary to evaluate its performance in more generic spatial networks. Its high sensitivity to natural structural features can be exploited to discover inherent hierarchical structures, if present. If this method is proven to be useful in those generic spatial networks, then its simplicity and self-containment will make it a good choice for first-level unsupervised learning techniques targeted at spatially embedded networks.

Conclusions

Here we have proposed a new modularity function that estimates the quality of partition of spatially embedded networks without employing a null model. Our new modularity function is inspired by the Hamiltonian of a spin-glass system. Notably, it incorporates geometric constraints through the use of a Heaviside function. Though arbitrary interactions may be incorporated, we employed couplings inspired from the null model (without explicitly using a null model). Our method accounts for the presence and absence (wherever applicable) of inter-community edges, which is neglected in nearly all kinds of formulations for modularity. The new modularity function is able to determine the “best” resolution scale at which the community structure has highest associability with the naturally identifiable structures. Our modularity function lucidly detected hierarchical structure present in the synthetically generated network with clearly its associated peaks and troughs as a function of the scale parameter; this was not the case for NG modularity. Taken together, our results suggest that the modified modularity function is more appropriate than the currently existing methods for unsupervised learning and analysis of spatially embedded networks.

References

1. Gastner, M. T. & Newman, M. E. J. The spatial structure of networks. *Eur. Phys. J. B* **49**(2), 247–252 (2006).
2. Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
3. Papadopoulos, L., Puckett, J. G., Daniels, K. E. & Bassett, D. S. Evolution of network architecture in a granular material under compression. *Phys. Rev. E* **94**(3), 032908 (2016).
4. Majumdar, T. S. & Behringer, R. P. Contact force measurements and stress-induced anisotropy in granular materials. *Nature* **435**, 1079 (2005).
5. Radjai, F., Wolf, D. E., Jean, M. & Moreau, J. J. Bimodal character of stress transmission in granular packings. *Phys. Rev. Lett.* **80**(1), 61 (1998).
6. Bassett, D. S., Owens, E. T., Daniels, K. E. & Porter, M. A. Influence of network topology on sound propagation in granular materials. *Phys. Rev. E* **86**(4), 041306 (2012).
7. Liu, C. H. Spatial patterns of sound propagation in sand. *Phys. Rev. B* **50**(2), 782 (1994).
8. Ronhovde, P. & Nussinov, Z. Local resolution-limit-free Potts model for community detection. *Phys. Rev. E* **81**(4), 046114 (2010).
9. Ronhovde, P. *et al.* Detecting hidden spatial and spatio-temporal structures in glasses and complex physical systems by multiresolution network clustering. *Eur. Phys. J. E* **34**(9), 105 (2011).

10. Good, B. H., de Montjoye, Y. A. & Clauset, A. Performance of modularity maximization in practical contexts. *Phys. Rev. E* **81**(4), 046106 (2010).
11. Duch, J. & Arenas, A. Community detection in complex networks using extremal optimization. *Phys. Rev. E* **72**(2), 027104 (2005).
12. Agarwal, G. & Kempe, D. Modularity-maximizing graph communities via mathematical programming. *Eur. Phys. J. B* **66**(3), 409–418 (2008).
13. Guimera, R. & Amaral, L. A. N. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
14. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**(2), 026113 (2004).
15. Newman, M. E. J. Fast algorithm for detecting community structure in networks. *Phys. Rev. E* **69**(6), 066133 (2004).
16. Mucha, P. J., Richardson, T., Macon, K., Porter, M. A. & Onnela, J. P. Community structure in time-dependent, multiscale, and multiplex networks. *Science* **328**, 876–878 (2010).
17. Bassett, D. S. *et al.* Dynamic reconfiguration of human brain networks during learning. *Proc. Nat. Acad. Sci.* **108**(18), 7641–7646 (2011).
18. Onnela, J. P. *et al.* Structure and tie strengths in mobile communication network. *s. Proc. Nat. Acad. Sci.* **104**(18), 7332–7336 (2007).
19. Kivela, M. *et al.* Multilayer networks. *Journal of complex networks* **2**(3), 203–271 (2014).
20. Boccaletti, S. *et al.* The structure and dynamics of multilayer networks. *Phys. Rep.* **544**(1), 1–122 (2014).
21. Expert, P., Evans, T. S., Blondel, V. D. & Lambiotte, R. Uncovering space-independent communities in spatial networks. *Proc. Nat. Acad. Sci.* **108**(19), 7663–7668 (2011).
22. Bassett, D. S. *et al.* Robust detection of dynamic community structure in networks. *Chaos* **23**(1), 013142 (2013).
23. Bassett, D. S., Owens, E. T., Porter, M. A., Manning, M. L. & Daniels, K. E. Extraction of force-chain network architecture in granular materials using community detection. *Soft Matter* **11**(14), 2731–2744 (2015).
24. Muzzio, F. J., Shinbrot, T. & Glasser, B. J. Powder technology in the pharmaceutical industry: the need to catch up fast. *Powder Technol.* **124**, 1–7 (2002).
25. Cleary, P. W. & Sawley, M. L. DEM modelling of industrial granular flows: 3D case studies and the effect of particle shape on hopper discharge. *Appl. Math. Model.* **26**(2), 89–111 (2002).
26. Cleary, P. W. DEM simulation of industrial particle flows: case studies of dragline excavators, mixing in tumblers and centrifugal mills. *Powder Technol.* **109**, 83–104 (2000).
27. Cundall, P. A. & Strack, O. D. L. A discrete numerical model for granular assemblies. *Geotechnique* **29**, 47–65 (1979).
28. Cundall, P. A. Formulation of a three-dimensional distinct element model—Part I. A scheme to detect and represent contacts in a system composed of many polyhedral blocks. *Int. J. Rock Mech. Min. Sci. & Geomech. Abstr.* **25**, 107–116 (1988).
29. Pöschel, T. & Schwager, T. *Computational Granular Dynamics 1st ed. Ch. 2*, 13–28 (Springer Science & Business Media, 2005).
30. Kishore, R., Das, S., Nussinov, Z. & Sahu, K. K. Kinetic instability, symmetry breaking and role of geometric constraints on the upper bounds of disorder in two dimensional packings. *Sci. Rep.* **6**, srep26968 (2016).
31. Kishore, R., Krishnan, R., Satpathy, M., Nussinov, Z. & Sahu, K. K. Abstraction of meso-scale network architecture in granular ensembles using ‘big data analytics’ tools. *J. Phys. Comm.* **2**(3), 031004 (2018).
32. Ronhovde, P. Physical models in community detection with applications to identifying structure in complex amorphous systems. *ProQuest Dissertations Publishing* (2010).
33. Reichardt, J. & Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* **74**(1), 016110 (2006).

Acknowledgements

K.K.S. was partially funded by UAY project (2017/84/IITBBS-006) and Z.N. was partially funded by NSF grant number 1411229.

Author Contributions

R.K., Z.N., K.K.S. conceived the idea, R.K. and A.K.G. performed all the simulations, all contributed equally in manuscript preparation.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-39180-8>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019