Abstract

Network partitioning, an unsupervised method of learning helps in discovering the ‘communities’ that reside in a network and share common properties of interest. We have partitioned a network generated from an ensemble of granular materials by using a highly accurate and resolution limit free community detection model proposed by Ronhovde and Nussinov (RN model). We associate the best communities with naturally identifiable structures in the ensemble to find out the ‘best’ resolution for partitioning the network. The Shannon entropy distribution of individual communities at different resolutions indicates existence of hierarchical structure in the network which is non-obvious and otherwise non-observable. Effect of disorder (irregular regions present in the granular ensemble) on the community evolution in the invariant network and its stability is analysed.

1. Introduction

Community detection is one of the most popular data mining techniques through graph partitioning because of its generality and hence it can be applied almost in every field including the Internet and its topology [1, 2], agriculture [3], finance [4], communication [5], sociology [6], biology [7–9] etc wherever data can be represented in the form of a network. Because of technological advancements, these fields create large amount of unprocessed data resulting in the ‘Big data problem’. Community detection can be used as a tool in finding useful information from a huge chunk of data. Graph analysis has become crucial in understanding the features of these systems. For instance, social network analysis started in the 1930’s and has become one of the most important topics in sociology [10, 11] and off-late, in the development of social-networking platforms that has dramatically augmented human-human communication paradigms. In biology, community detection can help in identifying cancer cells at early stages [12] and also help in confining the epidemic spreading by immunizing the highly connected infected persons/communities which are typically represented by nodes in a graph-architecture [13, 14]. Today, the leading internet based search engines are using community detection algorithms also, in addition to their core algorithm, to make search easier, faster and more relevant. In the past decade there has been a surge of interest in model development for detection of such communities in real world networks. Community detection in large networks can help to reveal relevant modules at meso-sopic scales that can affect or explain the global behavior of the system. This is of great value in understanding a complex network, particularly for isolating the sub-graphs that influence relevant network-dynamics more than others. There are several approaches to detect dense groups/communities in such networks [6, 15–22]. These approaches are reviewed in [23]. The accuracy of these approaches using benchmark networks where the ‘ground truth’ is known [24], can be found in [25–27]. The appropriateness of the outcomes of these community detection methods can be measured with the help of different metrics [28]. Efforts have been made in developing functions that evaluate the goodness of an identified community. These quality functions play important role in
evaluating the process. The present work uses one such method developed by Ronhovde and Nussinov [29, 30] (hereafter referred as RN model), based on spin-glass-type Potts model for identifying communities in granular ensembles. The RN model is a classic example of nature-inspired optimization technique, where fictitious spin states are imposed on a granular network (details of the method for construction of this granular network can be found in supplementary S1) is available online at stacks.iop.org/JPCO/2/031004/mmedia). In this study, only static (frozen in time, no ‘true’ dynamics involved) granular networks are considered. The evolution of the ‘imposed fictitious’ spin dynamics is observed on the otherwise invariant network. The imposition of the fictitious spin states allows the calculation of the energy (Hamiltonian) of the system. The Hamiltonian of the system (or network) attempts to model the spin-spin interaction similar to a magnetic material and ground state of the Hamiltonian corresponds to the optimal partitioning of the spin ‘degrees of freedom’ that manifests in the form of magnetic domains in the material. Though there can be exponentially large numbers of possible solutions (this is a NP-Complete problem [31]) corresponding to different domain structures, but most solutions are non-optimal and therefore of not prime importance. The optimal ‘communities’ are found by identifying the ground states of the spin-network and therefore, these ‘communities’ are analogous to the magnetic domains in the relaxed (annealed) state. Since the RN model uses a ‘local’ measure of community structure, this model is free from resolution limit [32] and one of the best algorithms currently available both in terms of speed and possible system size that can be handled [29]. In the present study, the community detection algorithm were applied on granular networks [33–37] to analyze its behavior under compression, mainly the formation of complex and collective structures, which can be best identified as ‘communities’. These structures influence bulk properties such as mechanical stability and acoustic transmission [38, 39], among others. Granular network are also important to understand the role of ordered granular chains, such as force chains, in supporting acoustic bands [40–42], heterogeneous pattern of force transmission [43]. In the present work, RN model is applied on a network generated from granular ensembles, which is described in detail in the next section.

2. Methods

2.1. Development of the granular network

The granular ensemble is generated by simulating a packing of 7278 macroscopic (radius 0.01 m) 2D disks by isotropic mechanical contraction (IMC) of the bounding circular wall at a fixed rate (0.5 m/s radially) using Discrete Element Method (DEM) [44–47]. In IMC, the packing happens purely because of the external constraints in complete absence of cohesive force. The particles in our study, driven by a bounding wall, progressively move towards the center of the container. In doing so, the particles lose energy during collisions among themselves and with the wall owing to friction and damping. To avoid excessive plastic deformation, the inward compressive motion of the external walls is terminated when overall area density (or, surface coverage) of ∼91% is achieved (at ∼3.18 s). While, the walls do not move beyond that instant of time, the system is further allowed to evolve by the virtue of its stored potential energy (relaxation). The simulations end after making certain that relaxation is achieved by waiting for a sufficiently long time (until 5.4 s) and no appreciable change in dynamical state is observed. The final packing is predominately hexagonal with minor but varied degrees of departures from six fold coordination (figure 1(a)). The network is generated from granular ensemble by considering edge between two granules (disks) if the distance between their center is less than or equal to sum of their radii (see S1). Due to such topological arrangement of disks, the network generated is a predominantly regular graph having regions of 6–regular sub-graphs separated by irregular graph boundaries (figure 1(b)). These boundaries play a vital role in clustering of such graphs because of the difference in degree distribution of these regions with those in regular regions.

The data in the form of files which contains network connection information (graph.gml) and coordinates of each node as well as degree of each node (Particle_pos_with_coord_no.xls) can be downloaded from https://sites.google.com/site/kksresearchgroup/j-phys-comm.

2.2. RN Model (Ronhovde and Nussinov [29, 30])

Performance of any graph partitioning model crucially depends on the choice of a quality function and its implementation. It should not only account for the connected edges but should consider the missing edges also. RN model directly penalizes for missing links within a community. This makes it highly accurate, a local model for general graphs (weighted, unweighted, and directed), and free from the resolution limit (see S2). The Potts Hamiltonian of a network (equation (1)) considers intra-community links as a favourable condition for a well-defined community structure and the opposite holds for missing edges inside communities. These arrangements lower the energy (Hamiltonian) of the system [48].
The un-weighted contact network is generated from granular ensemble by considering edge between two particles if the distance between their centers is less than or equal to sum of their radii) with un-weighted edges having 6-regular as well as irregular graph regions (shown in inset). RN model has access only to the connectivity matrix obtained from the network. It has no access to position of individual particles. For the sake of proper visualization, the particles are drawn using position information.

\[
H(\{\sigma\}) = -\frac{1}{2} \sum_{i,j} (a_{ij}A_{ij} - \gamma b_{ij}J_{ij}) \delta(\sigma_i, \sigma_j)
\]

where, \(H\) is the Hamiltonian of the system, \(a\) and \(b\) are the weights for connected and missing edges respectively, \(A\) is the adjacency matrix, \(J\) denotes the missing edges \((J = 1 - A)\), \(\delta\) refers to the Kronecker delta, \(\sigma\) refers to the spin (or community) type and the subscripts \(i\) and \(j\) denotes the indices of two arbitrary nodes chosen at random. Equation (1) can be applied on weighted (with weight \(a_{ij}\) for edge between node \(i\) and \(j\) and \(b_{ij}\) for missing link) as well as unweighted graphs (with \(a = 1\) and \(b = 1\)). Spins interact only with other spins in the same community (when \(\sigma_i = \sigma_j\)). The model weight \(\gamma\) is used to adjusts the energy trade-off between the two types of interactions i.e. attractive (connected) and repulsive (not connected). It allows model to adjust the resolution of the community solution and discussed in detail in the next section. The model weight \(\gamma\) used in equation (1) is related to the minimum edge density of each community as,

\[
p_{\text{min}} \geq \frac{\gamma}{\gamma + 1}
\]

3. Results

The un-weighted contact network is generated from granular ensemble by considering edge between two granules (disks) if the distance between their centers is less than or equal to sum of their radii (this is possible since the disks are deformable). Community detection in these networks is difficult because of relatively uniform degree distribution as compared to the distributions in scale free networks. This is particularly so because, at most occasions, the graph topology is relatively feature-free and general-purpose optimization techniques perform sub-optimally. It is of prime importance to note that the RN model had no access to the position of the individual particles directly and therefore it is impossible for this algorithm to extract any feature from geometrical patterns (see S2). In the present study, the RN model had to exclusively depend on the connectivity information, which, in a way, captures the neighbourhood of the particles. However as an aid to our analysis, the particles will be depicted by their coordinates so that visual correlation with RN model output can be performed.

A community evolved from the network is expected to have higher internal links as compared to external links. The in-versus out-bound edge ratio \((k_{in}/k_{out})\) is therefore, in general, more than unity. The granular network used here, is densely packed and therefore has 6-regular sub-graphs (most dense regions) separated by irregular graph boundaries (relatively less dense regions) as can be readily identified in figure 1(b).

These less dense irregular boundaries are likely to act as a template for separation between two communities, because in that case the value of \((k_{in}/k_{out})\) for the communities will increase leading to more accurate partitioning as the degree of nodes located in irregular regions are less when compared to nodes in regular regions. Therefore, in this particular work, the ‘grain boundary-like structures’ (or, alternatively, the deformed or amorphous regions) can be used as a visual tool to qualitatively evaluate this particular type of unsupervised form of learning (the RN model). The average edge ratio \((k_{in}/k_{out})\) for communities (averaged over all communities in the network) obtained at different resolutions monotonically increases with decrease in resolution (figure 2(a)) signifying an inverse relationship. The average community size is also inversely
proportional to resolution scale (figure 2(a)) and edge ratio is directly proportional to community size (figure 2(b)). These observations are mutually consistent and substantiate the fact that $\gamma$ can be used as a resolution scale (since community size is inversely proportional to $\gamma$, figure 2(a)) and a crucial parameter in identifying hierarchical structures.

The quality of partitioning for given granular network can be quantified by different parameters. Some of them are discussed below and results are presented for the same.

3.1. Shannon entropy (SE)

Entropy is a quantitative measure of irregularity. SE is calculated to measure the structural information content, in particular, the irregularity of the granular network and its partition obtained from RN model. If the probability distribution of node degree inside a given community is $P = (P_1, P_2, \ldots, P_d)$ where $d$ is the maximum node degree in the community, then the SE for a community $I_c$ is given as

$$I_c = -\sum_{i=1}^{d} p_i \log p_i$$

where $p_i = n_i / N$ is the probability of finding a node with degree $i$ in community $c$, $n_i$ is the number of nodes with degree $i$ in community $c$, $N$ is the total number of nodes in the network and $\log$ is with base 6 (maximum possible value of $d$ in the present work). SE will attain maximum value for a uniform degree distribution where $P = (1/d, 1/d, \ldots, 1/d)$, and minimum value of zero for a delta function, for example $P = (1, 0, \ldots, 0)$. The SE of network $I_N$ is the summation over entire communities.

$$I_N = -\sum_{j=1}^{q} \sum_{i=1}^{d} p_{ij} \log p_{ij} = -\sum_{j=1}^{q} I_{c,j}$$

where, $q$ is the number of communities in the network. The SE distribution of individual communities obtained from contact network at different resolutions indicates existence of hierarchical structure (figure 3). These structures can be used to extract the useful hidden meso-scale structural information from these granular ensembles. Such information will be crucial to understand the behaviour of these ensembles when the system responds to external stimuli, for example, an applied force. The variation in SE of smaller size communities is large and it converges as size of community increases. Zero SE for a community indicates that the degree of all the nodes in the community is same.

3.2. Newman modularity

Modularity ($Q$) is a measure of goodness or strength of partitioning of a network. It is the fraction of edges within communities minus the expected value of that fraction if the positions of the edges are randomized [49]. Newman [50] developed a quantitative measure of modularity by calculating the difference between the actual number of edges between any two nodes and the expected number of edges between them (equation (5)).
$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{d_i d_j}{2m} \right) \delta(\sigma_i, \sigma_j)$$

(5)

where, $m$, $d$ and $\sigma$ are total number of edges, degree of node and community membership. $A$ is the adjacency matrix. We have partitioned the granular network at different resolution scales. Modularity measurement of these partitioned networks helps in associating the best communities with the natural structures in the physical systems to find out the ‘best’ resolution for partitioning of such network. High value of modularity of any community indicates that it has high internal edge density and little edge connection with the other communities. Modularity of partitioned network first increases with decrease in resolution ($\gamma$ value) but after peaking at $\gamma = 0.002$ it decreases again (figure 4), suggests that this peak corresponds to the ‘best’ resolution scale for partitioning of given contact network as it creates most modular structures. This ‘best’ resolution scale is system dependent and it is generally linked to the structural features that might be present in the structure, which in turn, is sensitive to perturbations applied in the network.

The communities obtained at ‘best’ resolution scale are shown in figure 5(a). For proper visualization, the network is again converted into granular ensemble with particles center as node and same color is used for particles of same community. It is evident that most of the irregular region nodes fall in the boundary of two or
more communities. They act as connecting nodes between two communities. These nodes are termed as boundary nodes and nodes which connect internally are termed as inner nodes. The difference in average node degree (equation (6)) of these boundary and inner nodes for all the communities of 'best' clustered network have positive value which indicates that the boundary nodes are having lower average degree than the inner nodes (figure 5(d))

$$\Delta k = \frac{1}{N_b} \sum_{i=1}^{N_b} k_i - \frac{1}{N_i} \sum_{j=1}^{N_i} k_j$$ (6)

where, $N_b$ and $N_i$ is the number of inner and boundary nodes in the given community and $k$ is the degree of node. High positive value of $\Delta k$ represents a community with high internal edge density and less external links with other communities.

The irregular regions have profound influence on the stability (or robustness) of a community, which is observed by varying gamma value. Ideally the size of communities should reduce with increase in gamma value (figure 2(a)) but some communities do not follow this general trend. We have marked those communities (figure 6) and tried to find out possible reasons behind this unexpected observation.

If a community with regular graph region is surrounded by disordered regions (Ex. communities 1, 2 and 3 in figure 6)) then the community structure is robust with respect to a perturbation in gamma. In the absence of such clearly demarcated structure (Ex. Communities 22, 23 and 24 in figure 6)) these communities become highly sensitive to perturbations in gamma. The variation in size of community 1 with gamma is minimal as compared to community 22, 23 and 24. This indicates that these irregular boundaries act as barrier for growth of communities in the Hamiltonian spin dynamics as implemented in Pott’s model based RN method for community detection.
4. Conclusion

We have partitioned a granular network by using RN model, which is an unsupervised machine learning method. The ratio of in-versus out-bound edges ($k_{in}/k_{out}$) of a community and its size increases with decrease in resolution scale ($\gamma$). The Shannon entropy distribution of communities in a network at different resolutions indicates existence of hierarchical structures in the given network. The maximum of modularity is indicative of the best resolution scale for clustering. The communities of best partitioned network are associated with natural structures in the physical systems (granular ensemble). Communities completely enclosed by irregular graph nodes are less susceptible to perturbation in $\gamma$ since typically they are surrounded by irregular (amorphous) graph regions.

Acknowledgments

ZN was partially supported by the NSF under grant number DMR-1411229.

ORCID iDs

Kisor K Sahu @ https://orcid.org/0000-0003-2488-0420

References

[1] Danon L, Diaz-Guilera A and Arenas A 2006 The effect of size heterogeneity on community identification in complex networks J. Stat. Mech.: Theory Exp. 11 P11010.
[2] Lancichinetti A, Fortunato S and Kertész J 2009 Detecting the overlapping and hierarchical community structure in complex networks New J. Phys. 11 033015.
[3] Girvan M and Newman M E J 2002 Community structure in social and biological networks Proc. Natl Acad. Sci. USA 99 8271–6
[4] Palla G, Derényi I, Farkas I and Vicsek T 2005 Uncovering the overlapping community structure of complex networks in nature and society Nature (London) 435 814–8
[5] Clauset A 2005 Finding local community structure in networks Phys. Rev. E 72 026132.
[6] Blondel V D, Guillaume J L, Lambiotte R and Lefebvre E 2008 Fast unfolding of communities in large networks J. Stat. Mech.: Theory Exp. 10 10008

[7] Xu J and Chen H 2005 CrimeNet explorer: a framework for criminal network knowledge discovery ACM Trans. Inf. Syst. Secur. 23 201–26

[8] Masuda N 2009 Immunization of networks with community structure New J. Phys. 11 123018

[9] Guimerà R and Amaral L A N 2005 Functional cartography of complex metabolic networks Nature (London) 433 895–900

[10] Wasserman S and Faust K 1994 Social Network Analysis: Methods and Applications vol 8 (Cambridge, UK: Cambridge University Press)

[11] Scott J 2017 Social Network Analysis 4th edn (London, UK: SAGE Publication)

[12] Bontemps T E 2017 Network inference and community detection, based on covariance matrices, correlations and test statistics from arbitrary distributions Commun. Stat.-Theor. 46 9150–65

[13] Rosvall M, Esquivel A V, Lancichinetti A, West J D and Lambiotte R 2014 Memory in network flows and its effects on spreading dynamics and community detection Nat. Commun. 5 4630

[14] Pastor-Satorras R, Castellano C, Van Mieghem P and Vespignani A 2015 Epidemic processes in complex networks Rev. Mod. Phys. 87 925

[15] Lancichinetti A, Kivela M, Saramäki J and Fortunato S 2010 Characterizing the community structure of complex networks PLoS ONE 5 e11976

[16] Clauset A, Newman M E J and Moore C 2004 Finding community structure in very large networks Phys. Rev. E 70 066111

[17] Rosvall M and Bergstrom C T 2007 An information-theoretic framework for resolving community structure in complex networks Proc. Natl Acad. Sci. 104 7327–31

[18] Rosvall M, Axelsson D and Bergstrom C T 2009 The map equation Eur. Phys. J.-Spec. Top. 178 13–23

[19] Raghavan U N, Albert R and Kumara S 2007 Near linear time algorithm to detect community structures in large-scale networks Phys. Rev. E 76 036106

[20] Newman M E J 2006 Finding community structure in networks using the eigenvectors of matrices Phys. Rev. E 74 036104

[21] Reichardt J and Bornholdt S 2006 Statistical mechanics of community detection Phys. Rev. E 74 016110

[22] Pons P and Latapy M 2005 Computing communities in large networks using random walks Computer and Information Sciences-ICISIC 2005: 20th Int. Symp. (Istanbul, Turkey) pp 284–93 (Springer)

[23] Albert R and Barabási A L 2002 Statistical mechanics of complex networks Rev. Mod. Phys. 74 47–97

[24] Lancichinetti A, Fortunato S and Radicchi F 2008 Benchmark graphs for testing community detection algorithms Phys. Rev. E 78 046110

[25] A. Hric D, Darst R K and Fortunato S 2014 Community detection in networks: structural communities versus ground truth Phys. Rev. E 90 062805

[26] Zhang P 2015 Evaluating accuracy of community detection using the relative normalized mutual information J. Stat. Mech.-Theory E P11006

[27] Yang Z, Algesheimer R and Tesseme C J 2016 A comparative analysis of community detection algorithms on artificial networks Sci. Rep. 6 srep30750

[28] Chakraborty T, Dalmia A, Mukherjee A and Garguly N 2017 Metrics for community analysis: a survey, ACM Comput. Surv. 50 54

[29] Ronhovde P and Nussinov Z 2009 Multiresolution community detection for megascale networks by information-based replica correlations Phys. Rev. E 80 016109

[30] Ronhovde P and Nussinov Z 2010 Local resolution-limit-free Potts model for community detection Phys. Rev. E 81 046114

[31] Carey M R, Johnson D S and Stockmeyer L 1974 Some simplified NP-complete problems Proc. of the 6th Annual ACM Symp. on Theory of Computing pp 47–63 ACM

[32] Fortunato S and Barthelemy M 2007 Resolution limit in community detection P. Natl. A. Sci. India A 104 36–41

[33] Giusti C, Papadopoulos L, Owens E T, Daniels K E and Bassett D S 2016 Topological and geometric measurements of force-chain community structure Phys. Rev. E 94 032909

[34] Papadopoulos L, Puckett J G, Daniels K E and Bassett D S 2016 Evolution of network architecture in a granular material under compression Phys. Rev. E 94 032908

[35] Bassett D S, Owens E T, Porter M A, Manning M L and Daniels K E 2015 Extraction of force-chain network architecture in granular materials using community detection Soft Matter 11 2731–44

[36] Navakas R, Dziugys A and Peters B 2010 Application of graph community detection algorithms for identification of force clusters in squeezed granular packs Modern Build. Mater. Struct. Tech. 14 980–3

[37] Navakas R, Dziugys A and Peters B 2014 A community-detection based approach to identification of inhomogeneities in granular materials Physica A: Statistical Mechanics and its Applications 407 312–31

[38] Owens E T and Daniels K E 2011 Sound propagation and force chains in granular materials Europhys. Lett. 94 54005

[39] Hidalgo R C, Grosse C U, Kun F, Reinhardt H W and Herrmann H J 2002 Evolution of percolating force chains in compressed granular media Phys. Rev. Lett. 89 205501

[40] Nesterenko V F 1993 Propagation of nonlinear compression pulses in granular media J. Appl. Mech. and Tech. Phys. 24 733–43

[41] Coste C, Falcon E and Fauve S 1997 Solitary waves in a chain of beads under Hertz contact Phys. Rev. E 56 6104

[42] Job S, Melo F, Sokolow A and Sen S 2007 Solitary wave trains in granular chains: experiments, theory and simulations Granul. Matter. 10 13–20

[43] Daniels K E 2017 The role of force networks in granular materials EPJ Web of Conferences. EDP Sciences 140 01006

[44] Cundall P A and Strack O D L 1979 A discrete numerical model for granular assemblies Geotechnique 29 47–65

[45] Cundall P A 1988 Formulation of a three-dimensional distinct element model—I. A scheme to detect and represent contacts in a system composed of many polyhedral blocks Int. J. Rock Mech. Min. Sci. & Geomech. Abstr. 25 107–16

[46] Pöschel T and Schwager T 2005 ComputationalGranularDynamics1stedn(Berlin,Germany:SpringerScience&BusinessMedia)Ch2pp13–28

[47] Kishore R, Das S, Nussinov Z and Sahu K K 2016 Kinetic instability, symmetry breaking and role of geometric constraints on the upper bounds of disorder in two dimensional packings Sci. Rep. 6 srep26968

[48] Ronhovde P 2010 Physical Models in Community Detection with Applications to Identifying Structure in Complex Amorphous Systems Washington University in St. Louis

[49] Newman M E J 2004 Fast algorithm for detecting community structure in networks Phys. Rev. E 69 066133

[50] Newman M E J 2012 Communities, modules and large-scale structure in networks Nat. Phys. 8 25