New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

FAST TRACK COMMUNICATION • OPEN ACCESS

A nonuniform popularity-similarity optimization (nPSO) model to efficiently generate realistic complex networks with communities

To cite this article: Alessandro Muscoloni and Carlo Vittorio Cannistraci 2018 *New J. Phys.* **20** 052002

View the article online for updates and enhancements.

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics
Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

FAST TRACK COMMUNICATION

A nonuniform popularity-similarity optimization (nPSO) model to efficiently generate realistic complex networks with communities

Alessandro Muscoloni[1] and Carlo Vittorio Cannistraci[1,2]

[1] Biomedical Cybernetics Group, Biotechnology Center (BIOTEC), Center for Molecular and Cellular Bioengineering (CMCB), Center for Systems Biology Dresden (CSBD), Department of Physics, Technische Universität Dresden, Tatzberg 47/49, D-01307, Dresden, Germany
[2] Brain bio-inspired computing (BBC) lab, IRCCS Centro Neurolesi 'Bonino Pulejo', Messina, Italy

E-mail: kalokagathos.agon@gmail.com

OPEN ACCESS

RECEIVED
11 October 2017

REVISED
10 April 2018

ACCEPTED FOR PUBLICATION
26 April 2018

PUBLISHED
29 May 2018

Keywords: network models, hyperbolic geometry, community structure

Supplementary material for this article is available online

Abstract

The investigation of the hidden metric space behind complex network topologies is a fervid topic in current network science and the hyperbolic space is one of the most studied, because it seems associated to the structural organization of many real complex systems. The popularity-similarity-optimization (PSO) model simulates how random geometric graphs grow in the hyperbolic space, generating realistic networks with clustering, small-worldness, scale-freeness and rich-clubness. However, it misses to reproduce an important feature of real complex networks, which is the community organization. The geometrical-preferential-attachment (GPA) model was recently developed in order to confer to the PSO also a soft community structure, which is obtained by forcing different angular regions of the hyperbolic disk to have a variable level of attractiveness. However, the number and size of the communities cannot be explicitly controlled in the GPA, which is a clear limitation for real applications. Here, we introduce the nonuniform PSO (nPSO) model. Differently from GPA, the nPSO generates synthetic networks in the hyperbolic space where heterogeneous angular node attractiveness is forced by sampling the angular coordinates from a tailored nonuniform probability distribution (for instance a mixture of Gaussians). The nPSO differs from GPA in other three aspects: it allows one to explicitly fix the number and size of communities; it allows one to tune their mixing property by means of the network temperature; it is efficient to generate networks with high clustering. Several tests on the detectability of the community structure in nPSO synthetic networks and wide investigations on their structural properties confirm that the nPSO is a valid and efficient model to generate realistic complex networks with communities.

Introduction

In recent years the study of hidden geometrical spaces behind complex network topologies has led to several developments and, currently, the hyperbolic space seems to be one of the most appropriate in order to explain many of the structural features observed in real networks [1–12]. In 2012 Papadopoulos *et al* [5] introduced the popularity-similarity-optimization (PSO) model in order to describe how random geometric graphs grow in the hyperbolic space optimizing a trade-off between popularity and similarity. In this framework, the popularity of the nodes is represented by the radial coordinate in the hyperbolic disk, whereas the angular coordinates distance is the geometrical counterpart of the similarity between the nodes. Networks generated through the PSO model exhibit strong clustering and a scale-free degree distribution, two among the peculiar properties that usually characterize real-world topologies [13–15]. However, another important feature commonly observed is the community structure [16–18], which is lacking in the PSO model. The reason is that the nodes are arranged over the angular coordinate space according to a uniform distribution, therefore, since the connection probability is a decreasing function of the hyperbolic distance, there are not angular regions containing a cluster of spatially close nodes that are more densely connected between each other than with the rest of the network. This issue has been addressed in a following study by Zuev *et al* [19], introducing the geometric-preferential-attachment (GPA). The GPA couples the latent hyperbolic network geometry with preferential attachment of nodes to this geometry in order to generate networks with strong clustering, scale-free degree distribution and a non-trivial community structure [19]. The main assumption of the GPA model and simultaneously the main innovation with respect to the PSO model is that the angular coordinate space is not equally attractive everywhere. Practically, the GPA is characterized by heterogeneous angular attractiveness: regions of different attractiveness are designed according to the rationale that the higher the attractiveness of a region the higher the probability that the nodes are placed in that angular section. Although this general idea can be implemented in several ways, a high-level description of the procedure presented in the study of Zuev *et al* [19] is as follows (see Methods for details). For each new node entering in the network, a set of candidate positions is defined (angular coordinate sampled uniformly at random, radial coordinate mathematically fixed) and to every candidate position is assigned a probability depending on the number of nodes that would be 'close' to the entering node if it were placed in that position. The probability is also function of a parameter of initial attractiveness, which can be used to tune the heterogeneity of the angular coordinate distribution. However, the GPA model does not allow—at least in the form in which it is currently proposed—to directly control in an *explicit* and *efficient* way the number and size of the communities, a property that instead might be interesting, for example, while proposing a community detection benchmark. Furthermore, the GPA model does not take into account the possibility to vary the network temperature. For these reasons we here introduce a variation of the PSO model, which we call nonuniform PSO (nPSO) model, whose key aspects are the possibility of: (a) fixing the number and size of communities; (b) tuning their mixing property through the network temperature; (c) efficiently producing also highly clustered realistic networks. Finally, although we present the nPSO as a generative model for non-overlapping communities, we will discuss a strategy for taking into account also the presence of overlapping communities.

Methods

PSO model

The PSO model [5] is a generative network model recently introduced in order to describe how random geometric graphs grow in the hyperbolic space. In this model the networks evolve optimizing a trade-off between node popularity, abstracted by the radial coordinate, and similarity, represented by the angular coordinate distance, and they exhibit many common structural and dynamical characteristics of real networks.

The model has five input parameters:

- $N > 0$, number of nodes in the network;
- $m > 0$, equal to half of the average node degree;
- $T \geqslant 0$, network temperature, which controls the network clustering; the network clustering is maximized at $T = 0$, it decreases almost linearly for $T = [0,\ 1)$ and it becomes asymptotically zero if $T > 1$;
- $\beta \in (0,\ 1]$, popularity fading parameter, or alternatively $\gamma \geqslant 2$, exponent of the power-law degree distribution, due to the relationship $\gamma = 1\ +\ 1/\beta$;
- $\zeta = \sqrt{-K}\ > 0$, where K is the curvature of the hyperbolic plane. Since changing ζ rescales the node radial coordinates and this does not affect the topological properties of network [5], in the rest of the article we will consider $K = -1$.

Building a network in the hyperbolic disk requires the following steps:

(1) Initially the network is empty;

(2) At time $i = 1,\ 2,\ \ldots,\ N$ a new node i appears with radial coordinate $r_i = 2\ln(i)$ and angular coordinate θ_i uniformly sampled in $[0,\ 2\pi]$; all the existing nodes $j < i$ increase their radial coordinates according to $r_j(i) = \beta r_j + (1 - \beta)r_i$ in order to simulate popularity fading;

(3) If $T = 0$, the new node connects to the m hyperbolically closest nodes; if $T > 0$, the new node picks a randomly chosen existing node $j < i$ and, given that it is not already connected to it, it connects to it with probability

$$p(i,j) = \frac{1}{1 + \exp\left(\frac{h_{ij} - R_i}{2T}\right)} \tag{1}$$

repeating the procedure until it becomes connected to m nodes.
Note that

$$R_i = r_i - 2\ln\left[\frac{2T(1 - e^{-(1-\beta)\ln(i)})}{\sin(T\pi)m(1-\beta)}\right] \tag{2}$$

is the current radius of the hyperbolic disk, and

$$h_{ij} = \operatorname{arccosh}(\cosh r_i \cosh r_j - \sinh r_i \sinh r_j \cos\theta_{ij}) \tag{3}$$

is the hyperbolic distance between node i and node j, where

$$\theta_{ij} = \pi - |\pi - |\theta_i - \theta_j|| \tag{4}$$

is the angle between these nodes.

(4) The growing process stops when N nodes have been introduced.

GPA model

The GPA model is a variation of the original PSO model that couples the latent hyperbolic network geometry with preferential attachment of nodes to this geometry in order to generate networks with strong clustering, scale-free degree distribution and a non-trivial community structure [19].

The procedure to generate a network of N nodes is the same described in the previous section for the PSO model, with the main difference that the angular coordinate θ_i of the new node i is assigned as follows:

(a) Sample $\varphi_1, \ldots, \varphi_i$ in $[0, 2\pi]$ uniformly at random. The points (r_i, φ_j) for $j = 1 \ldots i$ represent candidate positions for the node.

(b) Define for each candidate position (r_i, φ_j) the attractiveness $A_i(\varphi_j)$ equal to the number of existing nodes that lie within hyperbolic distance r_i from it.

(c) Set the angular coordinate $\theta_i = \varphi_j$ with probability:

$$\Pi_i(\varphi_j) = \frac{A_i(\varphi_j) + \Lambda}{\sum_{k=1}^{i} A_i(\varphi_k) + \Lambda}, \tag{5}$$

where $\Lambda \geqslant 0$ is a parameter representing the initial attractiveness.

Note that the GPA model has been presented in the related study with only three input parameters, m, β and Λ, with the additional parameters of the PSO model considered in the setting $T = 0$ and $K = -1$.

nPSO model

The nPSO model is a variation of the PSO model introduced in order to confer to the generated networks an adequate community structure, which is lacking in the original model. Since the connection probability is a decreasing function of the hyperbolic distance, a uniform distribution of the nodes over the hyperbolic disk does not create agglomerates of nodes that are concentrated on angular sectors and that are more densely connected between each other than with the rest of the network. A nonuniform distribution, instead, allows one to do it by generating heterogeneity in the angular node arrangement. Given the parameters of the PSO model (N, m, T, γ) and a nonuniform probability distribution defined in $[0, 2\pi[$, the procedure to generate a network is the same described in the section for the uniform case, with the only difference that the angular coordinates of the nodes are not sampled uniformly but according to the given nonuniform probability distribution.

In particular, without loss of generality, we will concentrate on the mixture of distributions where the components are either Gaussian or Gamma distributions (figure 1(A)), which we consider suitable for describing how to build a nonuniform distributed sample of nodes along the angular coordinates of a hyperbolic disk, with communities that emerge in correspondence of the different components. For instance, given a Gaussian mixture distribution the communities will emerge in correspondence of the different Gaussians. In particular, a Gaussian mixture distribution is characterized by the following parameters [20]:

- $C > 0$, which is the number of components, each one representative of a community;
- $\mu_{1\ldots C} \in [0, 2\pi[$, which are the means of the components, representing the central locations of the communities in the angular space;

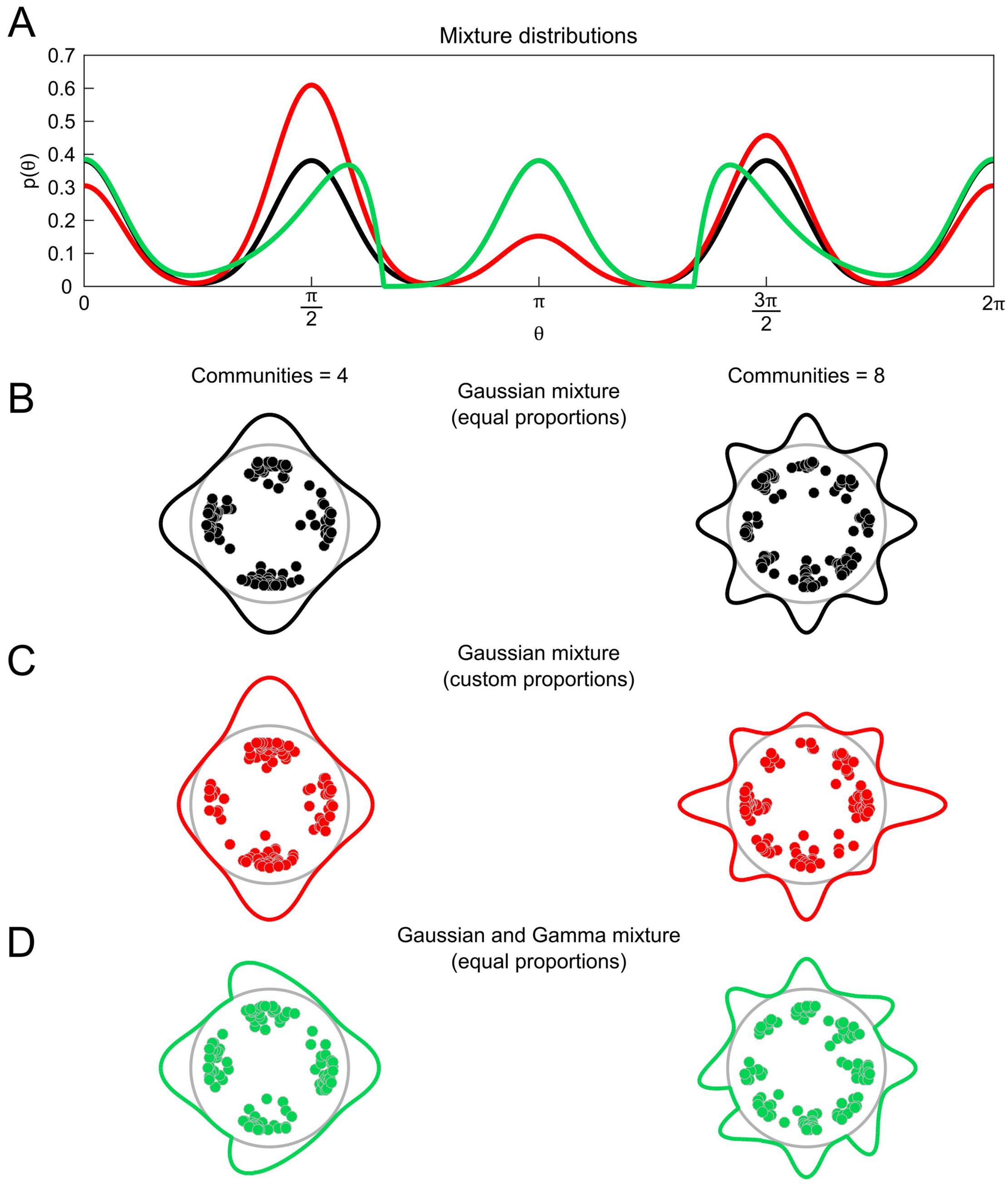


Figure 1. Nonuniform distributions of angular coordinates. (A) Plot of three different kinds of nonuniform distributions used for sampling the angular coordinates of the nodes: Gaussian mixture with equal proportions (nPSO1, black line), Gaussian mixture with custom proportions (nPSO2, red line), Gaussian and Gamma mixture with equal proportions (nPSO3, green line). The mixture distributions have 4 components, placing the mean of the components equidistantly over the angular space. More details on the parameters of the mixture distributions are in the Methods. (B)–(D) The same three kinds of mixture distributions in (A) are represented along the angular space of the hyperbolic disk, using 4 and 8 components. For each of them, a synthetic network has been generated using the nPSO model with parameters $N = 100, m = 2, T = 0, \gamma = 3$ and angular coordinates sampled according to the mixture distribution. The coordinates of the nodes are represented in the hyperbolic disk in order to provide an example of the emerging community structure.

- $\sigma_{1...C} > 0$, which are the standard deviations of the components, determining how much the communities are spread in the angular space; a low value leads to isolated communities, a high value makes the adjacent communities to overlap;

- $\rho_{1...C}(\sum_i \rho_i = 1)$, which are the mixing proportions of the components, determining the relative sizes of the communities.

Note that, although the means of the components are located in $[0, 2\pi[$, the sampling of the angular coordinate θ can fall out of this range. In this case, it has to be shifted within the original range using the modulo operator: $\theta = \text{modulo}(\theta, 2\pi)$.

Although the parameters of the Gaussian mixture distribution allow for the investigation of disparate scenarios, as a first case of study (figure 1(B)) we focused on the most straightforward setting. For a given number of components C, we considered their means equidistantly arranged over the angular space, the same standard deviation and equal mixing proportions:

- $\mu_i = \frac{2\pi}{C} * (i - 1) \quad i = 1 \ldots C$
- $\sigma_1 = \sigma_2 = \ldots = \sigma_C = \sigma$
- $\rho_1 = \rho_2 = \ldots = \rho_C = \frac{1}{C}.$

In particular, in our simulations we fixed the standard deviation to 1/6 of the distance between two adjacent means $\left(\sigma = \frac{1}{6} * \frac{2\pi}{C}\right)$, which allowed for a reasonable isolation of the communities independently from their number.

In a second scenario (figure 1(C)), we introduced asymmetries in the distribution of the nodes over the circumference by generating communities of different sizes, implemented using diverse mixing proportions for the components. In particular, in our simulations the mixing proportions have been randomly assigned.

As a last scenario (figure 1(D)), we considered a mixture of Gaussian and Gamma distributions, which is also characterized by asymmetries due to the presence of the Gamma components. Since this is not a mixture distribution as ordinary as the Gaussian one, in supplementary information (available online at stacks.iop.org/NJP/20/052002/mmedia) we will provide the details about how it has been built for our simulations. In all the scenarios, the community memberships are assigned considering for each node the component whose mean is at the lowest angular distance.

Computational implementations of the generative model algorithm

In the PSO model section, at step (3) of the generative procedure, it is presented how the new node establishes connections to m of the existing nodes. In particular, if $T > 0$, the new node i picks a randomly chosen existing node $j < i$ and, given that it is not already connected to it, it connects with probability $p(i, j)$, repeating the procedure until it becomes connected to m nodes. An example of pseudocode is:

```
targets = [1…i-1]
c = 0
while c < m
  j = random node uniformly sampled from targets
  rand_p = random number in [0, 1]
  if p(i, j) > rand_p
    add link from i to j
    remove j from targets
    c = c + 1;
  end
end
```

At the implementation level, the basic solution in MATLAB code would be:

```
targets = 1:(i-1);
c = 0;
while c < m
  idx = randi(length(targets));
  j = targets(idx);
  rand_p = rand(1);
  if p(i, j) > rand_p
    x(i, j) = 1;
    c = c + 1;
    targets(idx) = [];
  end
end
```

where x is the adjacency matrix of the network. We will refer to this as implementation 1.

As it will be commented in the Results and Discussion, this implementation has issues of time performance in specific cases. In fact, it is possible to note from equation (1) that the connection probability $p(i,\ j)$ decreases both for increasing hyperbolic distance and for decreasing temperature (when $h_{ij} > R_i$, which is true in the majority of the cases as shown in supplementary table 1, in particular for increasing network size). Therefore, while generating a network with low temperature and where many nodes are at high hyperbolic distance (for example sampling the angular coordinates from a Gaussian mixture distribution with 4 communities), most of the connection probabilities to the targets will be low. As a consequence, the *if* statement will result *false* in many iterations and the *while* loop will require a relevant computational time before that m connections are successfully established.

In order to solve this issue, we note that at each *while* loop iteration the connection probabilities to the target nodes (excluded the ones already connected) do not always cover the full range [0, 1]. In particular, at each iteration the maximum of these probabilities $\text{max\_}p = \max_{t \in \text{targets}} p(i,\ t)$ will be usually lower than 1. Since it is known *a priori* that any random sampling $\text{rand\_}p > \text{max\_}p$ will necessarily bring to a rejection of the connection independently from the target node chosen, the sampling range of the random probability rand_*p* can be restricted to [0, max_*p*]. In the critical conditions previously mentioned, where most of the connection probabilities are low, this adjustment can bring to a considerable speedup without biasing the link generation procedure. An example of pseudocode is:

```
targets = [1…i-1]
c = 0
max_p = $\max_{t \in \text{targets}} p(i,\ t)$
while c < m
  j = random node uniformly sampled from targets
  rand_p = random number in [0, max_p]
  if p(i, j) > rand_p
    add link from i to j
    remove j from targets
    max_p = $\max_{t \in \text{targets}} p(i,\ t)$
    c = c + 1;
  end
end
```

In case a programming language optimized for vector operations (i.e. MATLAB) is used, since vector operations are faster than loop-based operations, at each iteration m attempts of connection to target nodes can be done at once, reducing the number of iterations required to successfully establish m connections. Note that, while this adjustment is convenient only at the implementation level when using a programming language optimized for vectorization, the restriction of the probability sampling to the range [0, max_*p*] is valid in general.

The MATLAB code of the implementation would be:

```
targets = 1:(i-1);
c = 0;
max_p = max(p(i, targets));
while c < m
  if length(targets) > m
    idx = randsample(length(targets), m);
  else
    idx = 1:length(targets);
  end
  rand_p = rand(1, length(idx)) * max_p;
  idx = idx(p(i, targets(idx)) > rand_p);
  if ~is empty(idx)
    if length(idx) > m - c
      idx = randsample(idx, m - c);
    end
    x(i, targets(idx)) = 1;
    targets(idx) = [];
    max_p = max(p(i, targets));
    c = c + length(idx);
  end
end
```

We will refer to this as implementation 2.

A further variant that we propose is to sample the target nodes according to the theoretical probabilities $p(i,\ j)$. This solution ensures that at every iteration new connections are successfully established, avoiding rejections and making the procedure faster. An example of pseudocode is:

```
targets = [1…i-1]
for c = 1…m
    j = random node t sampled from targets with probabilities $\frac{\mathrm{p(i, t)}}{\sum_{u\,\in\,\mathrm{targets}} \mathrm{p(i,\ u)}}$
    add link from i to j
    remove j from targets
end
```

Given normalized connection probabilities from node i to U targets, $w(i,\quad t) = \frac{p(i, t)}{\sum_{u\,\in\,\mathrm{targets}} p(i,\ u)}$, the nonuniform sampling can be performed in the following way:

(a) Partition the interval [0, 1] in U subintervals $I(t)$ of sizes $w(i,\quad t)$

(b) Generate a random number $r \in [0,\quad 1]$

(c) The sampled target is the t such that $r \in I(t)$.

The MATLAB code of the implementation would be:

```
targets = 1:(i-1);
idx = datasample(targets, m, 'Replace', false, 'Weights', p(i,targets));
x(i,idx) = 1;
```

We will refer to this as implementation 3.

Note that, as for the previous implementation, the sampling of m targets at once is an adjustment convenient only at the implementation level when using a programming language optimized for vectorization, whereas the idea of sampling according to the theoretical probabilities is valid in general.

The computational complexity of the model using the three implementations is discussed in the next section, whereas their running time as well as the equivalence of the generated synthetic networks is discussed in the Results and Discussion.

Computational complexity

The generative procedure of the nPSO model mainly consists in a loop of N iterations, where for each iteration i a new node appears and connects to m of the existing nodes.

Let us firstly consider the degenerate case in which $m \approx N$. For approximately all the N iterations the connections m to establish are more than the existing nodes, therefore the new node i will simply connect to all the previous $i - 1$ nodes, with $O(i)$ operations. The computational complexity is given by:

$$\sum_{i=1}^{N} i = \frac{N \cdot (N+1)}{2} = O(N^2).$$

Let us now consider the more realistic case in which $m \ll N$. For approximately all the N iterations the connections m to establish are less than the existing nodes, therefore the new node i will connect to only m of them, and the time-dominant operations to create the links change depending on the temperature (whether zero or positive) and on which of the three implementations is adopted.

For $T = 0$, the connections are established with the m hyperbolically closest nodes, independently from the implementation. For each iteration i, $i - 1$ hyperbolic distances have to be computed in $O(i)$ operations, and then the m smallest ones have to be found, which can be obtained building a min-heap in $O(i)$ and retrieving the minimum m times in $O(m \log i)$. Considering $\sum_{i=1}^{N} \log i = O(N \log N)$ and $E = mN$, the computational complexity is given by:

$$\sum_{i=1}^{N}(i + m \log i) = \sum_{i=1}^{N} i + m\sum_{i=1}^{N} \log i = O(N^2) + m \cdot O(N \log N) = O(N^2 + E \log N).$$

For $T > 0$, for each iteration i the m links are instead established according to the connection probabilities $p(i, j)$. We will now analyse the three different implementations.

Using the implementation 1, for each link to create, one target node t is uniformly sampled and the connection is established with probability $p(i, t)$, otherwise rejected. Therefore a connection attempt costs constant time $O(1)$. The average connection probability to the targets changes over the iterations i and over the m links depending on the set of targets. Let us indicate with $\tilde{p}_1$ the average connection probability to the targets over the entire generative procedure using implementation 1. For each iteration i, on average $\frac{m}{\tilde{p}_1}$ connection attempts of cost $O(1)$ are performed and therefore at most $O\left(\frac{m}{\tilde{p}_1}\right)$ operations are required. The computational complexity is given by:

$$\sum_{i=1}^{N} \frac{m}{\tilde{p}_1} = O\left(N\frac{m}{\tilde{p}_1}\right) = O\left(E\frac{1}{\tilde{p}_1}\right).$$

Using the implementation 2, for each link to create, one target node t is uniformly sampled and the connection is established with probability $p(i, t)/\text{max\_p}$, otherwise rejected (max_p is the maximum probability between the targets). Computing the maximum costs $O(i)$ and since it has to be updated every time a connection is successfully established, for each iteration i its overall cost is $O(m \cdot i)$. Analogously to implementation 1, on average $\frac{m}{\tilde{p}_2}$ connection attempts of cost $O(1)$ are performed, where $\tilde{p}_2$ is the average connection probability to the targets over the entire generative procedure using implementation 2. The computational complexity is given by:

$$\sum_{i=1}^{N}\left(m \cdot i + \frac{m}{\tilde{p}_2}\right) = O(mN^2) + O\left(N\frac{m}{\tilde{p}_2}\right) = O\left(EN + E\frac{1}{\tilde{p}_2}\right) = O\left(E\left(N + \frac{1}{\tilde{p}_2}\right)\right).$$

Using the implementation 3, for each link to create, one target node t is nonuniformly sampled with probabilities $w(i, t) = \frac{p(i, t)}{\sum_{u \in \text{targets}} p(i, u)}$ and the connection is successfully established. The computation of the normalized probabilities and the nonuniform sampling have a cost of $O(i)$, which is performed exactly m times. The computational complexity is given by:

$$\sum_{i=1}^{N} m \cdot i = O(mN^2) = O(EN).$$

Note that the factor $\frac{1}{\tilde{p}_1}$ (in implementation 1) is expected to be higher with respect to $\frac{1}{\tilde{p}_2}$ (implementation 2) and they both mainly increase for low temperatures and when the hyperbolic distances are overall high (i.e. low number of communities). Therefore, depending on the model's parameter combination, the factors $\frac{1}{\tilde{p}_1}$ and $\frac{1}{\tilde{p}_2}$ can have a significant impact on the computational time, which will be discussed in the next section.

Results and discussion

The idea behind the nPSO is quite intuitive. The sampling of the angular coordinates from a uniform distribution—which is used by the standard PSO—can be generalized to sampling from any distribution with a desired shape. In particular, a nonuniform distribution would indicate the presence of regions with different levels of node attractiveness. In this study, without loss of generality, we will concentrate on the mixture of distributions where the components can be either Gaussian or Gamma distributions, which we consider suitable for describing how to build a nonuniform distributed sample of nodes along the angular coordinates of a hyperbolic disk, with communities that emerge in correspondence of the different distribution components. However, we want to stress that our nPSO model is general and can be implemented considering any mixture of desired distributions from which to sample the angular coordinates of the nodes.

Although the parameters of the Gaussian and Gamma mixture distributions built on the angular coordinate space allow for the investigation of disparate scenarios, in this work we focused on three straightforward settings, which are illustrated in figure 1. For a given number of communities C, in the first scenario (figure 1(B)), we consider a Gaussian mixture distribution of C components with the means equidistantly arranged over the angular space, the same standard deviation and equal mixing proportions (see Methods for details). In a second scenario (figure 1(C)), we introduced asymmetries in the distribution of the nodes over the circumference by generating communities of different sizes, implemented using diverse mixing proportions for the components. As a third and last scenario (figure 1(D)), we considered a mixture of Gaussian and Gamma distributions, which is also characterized by asymmetries due to the presence of the Gamma components. In all the scenarios, the community memberships are assigned considering for each node the component whose mean is at the lowest

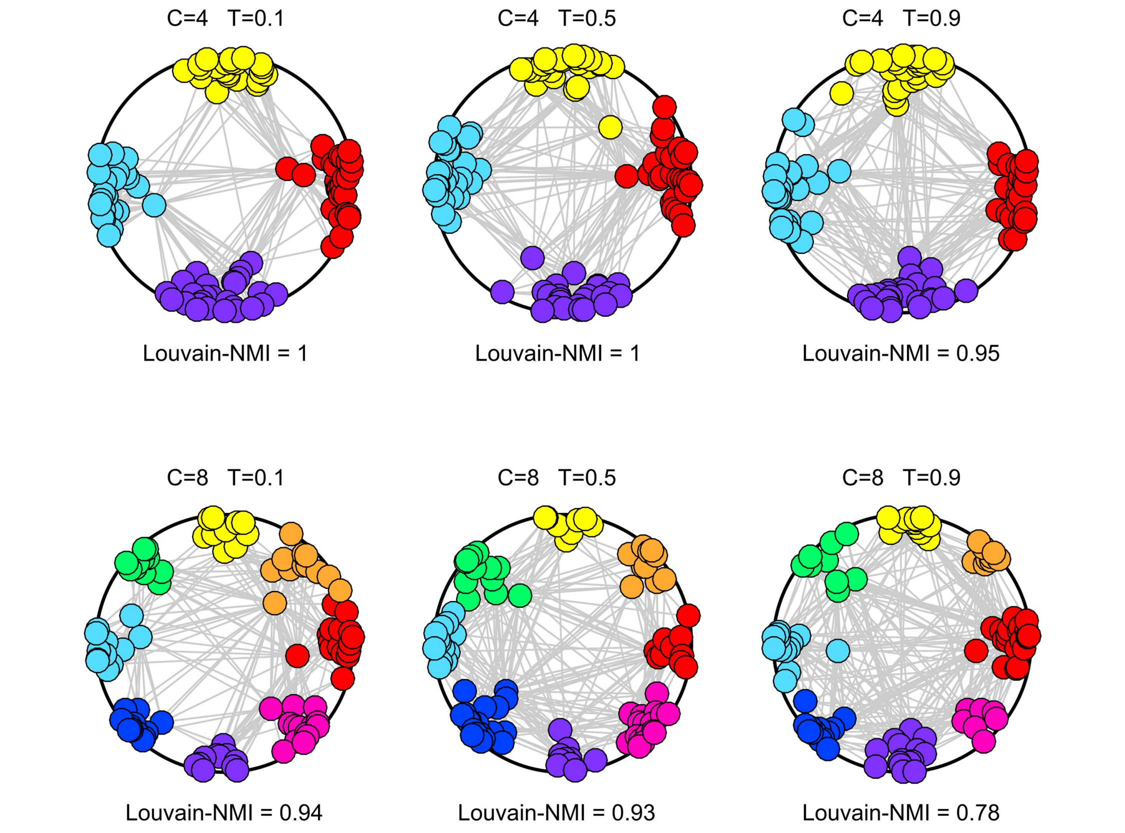


Figure 2. Communities generated using the nPSO model. Synthetic networks have been generated using the nPSO model with parameters $N = 100, m = 5, T = [0.1, 0.5, 0.9], \gamma = 3$ and angular coordinates sampled according to a Gaussian mixture distribution with equal proportions and components $C = [4, 8]$ (communities). For each combination of parameters, 10 networks have been generated. For each network the Louvain community detection method has been executed and the communities detected have been compared to the annotated ones computing the normalized mutual information (NMI). The plots show for each parameter combination a representation in the hyperbolic space of the network that obtained the highest NMI, whose value is reported. The nodes are coloured according to the communities as generated by the nPSO model.

angular distance. Figure 2 shows examples of networks in the hyperbolic space generated using the nPSO model (first scenario, Gaussian mixture distribution with equal proportions is considered) for different values of clustering (temperature, $T = [0.1, 0.5, 0.9]$) and community number ($C = [4, 8]$), while keeping the other parameters fixed ($N = 100, m = 5, \gamma = 3$). The related communities are also highlighted using different node colours. The figure indicates below each network also the normalized mutual information (NMI) [21], a measure of performance for evaluating the community detection, computed by comparing the nPSO ground-truth communities and the ones detected by Louvain [22], which is one of the state-of-the-art community detection algorithms [23] (see supplementary methods for details). We notice that the communities are perfectly detected both for $C = 4$ and $C = 8$ at low temperature, suggesting that a meaningful community structure is generated by the proposed model. For the same number of communities, if the temperature is increased the performance slightly decreases, because more inter-community links are established in the network, causing as expected a higher rate of wrong assignments by the community detection algorithm.

The next sections will be organized as follows: at first we will prove the equivalence of the three implementations for the generative model algorithm and we will discuss their computational efficiency; later—using the fastest implementation (which is the implementation 3) to generate numerous networks over diverse parameter combinations—we will propose a wide investigation on the detectability of the communities and on the topological properties of the synthetic networks generated by the nPSO.

Equivalence of the three implementations for link generation

Let us consider a node *i* that has to establish a connection with one over *U* target nodes.

Implementation 1. One target node *t* is chosen uniformly at random and a connection is established with probability $p(i, \ t)$.

Let us call C_t the event: node i connects with target t. The probability of the event is:

$$P(C_t) = \frac{1}{U}\cdot p(i,\ t).$$

Let us call C the event: node i connects with any of the targets. Taking into account that the event C is the union of the events $C_1,\ C_2,\ \ldots,\ C_U$ and that these events are mutually exclusive, the probability of the event is:

$$P(C) = P\left(\bigcup_{t\in U} C_t\right) = \sum_{t\in U} P(C_t) = \sum_{t\in U}\frac{1}{U}\cdot p(i,\ t) = \frac{1}{U}\cdot\sum_{t\in U} p(i,\ t).$$

In case the connection is rejected, another attempt is iteratively made until node i connects with any of the targets. In other words, the procedure is repeated until the event C occurs.

Therefore, the probability to eventually recruit the target t as a neighbour is given by the conditional probability that node i has connected with target t, given that event C occurred:

$$P(C_t|C) = \frac{P(C_t\cap C)}{P(C)} = \frac{P(C_t)}{P(C)} = \frac{\frac{1}{U}\cdot p(i,\ t)}{\frac{1}{U}\cdot\sum_{u\in U}p(i,\ u)} = \frac{p(i,\ t)}{\sum_{u\in U}p(i,\ u)}.$$

Implementation 2. One target node t is chosen uniformly at random and a connection is established with probability $\frac{p(i,t)}{\max_{u\in U}p(i,u)}$.

Following the same procedure as the implementation 1, we obtain:

$$P(C_t) = \frac{1}{U}\cdot\frac{p(i,\ t)}{\max\limits_{u\in U} p(i,\ u)} = \frac{1}{U}\cdot\frac{1}{\max\limits_{u\in U} p(i,\ u)}\cdot p(i,\ t),$$

$$P(C) = \frac{1}{U}\cdot\sum_{t\in U}\frac{p(i,\ t)}{\max\limits_{u\in U} p(i,\ u)} = \frac{1}{U}\cdot\frac{1}{\max\limits_{u\in U} p(i,\ u)}\cdot\sum_{t\in U} p(i,\ t),$$

$$P(C_t|C) = \frac{P(C_t)}{P(C)} = \frac{\frac{1}{U}\cdot\frac{1}{\max\limits_{u\in U}p(i,u)}\cdot p(i,\ t)}{\frac{1}{U}\cdot\frac{1}{\max\limits_{u\in U}p(i,u)}\cdot\sum_{u\in U}p(i,\ u)} = \frac{p(i,\ t)}{\sum_{u\in U}p(i,\ u)}.$$

Implementation 3. one target node t is chosen nonuniformly at random with probability $\frac{p(i,t)}{\sum_{u\in U}p(i,u)}$ and a connection is established.

Note that in this procedure the connection is never rejected. Therefore we obtain:

$$P(C_t) = \frac{p(i,\ t)}{\sum_{u\in U}p(i,\ u)},$$
$$P(C) = 1,$$
$$P(C_t|C) = \frac{P(C_t)}{P(C)} = \frac{p(i,\ t)}{\sum_{u\in U}p(i,\ u)}.$$

Since the probability $P(C_t|C)$ to eventually recruit the target t as a neighbour is the same for the three implementations, their equivalence is proven.

However, as a further demonstration that the generative procedure is not biased toward networks with different properties, in supplementary tables 2–6 we report for each PSO and nPSO parameter combination some of the main deterministic (in the sense that the measure does not depend by the stochastic generation of null models) topological measures of the networks generated using the three different implementations: clustering coefficient, characteristic path length, assortativity, local-community-paradigm (LCP) correlation and power-law exponent. The results confirm that these structural properties of the model are well preserved and there are not noteworthy changes introduced by the adoption of the algorithmic variants.

nPSO algorithm efficiency in generating networks with high clustering

One of the main drawbacks to use the original algorithmic implementation to establish links adopted by the PSO and GPA models also for the nPSO model, is the lack of efficiency in generating networks with communities characterized by high clustering (low temperature), when $T > 0$. As reported in figure 3 and supplementary

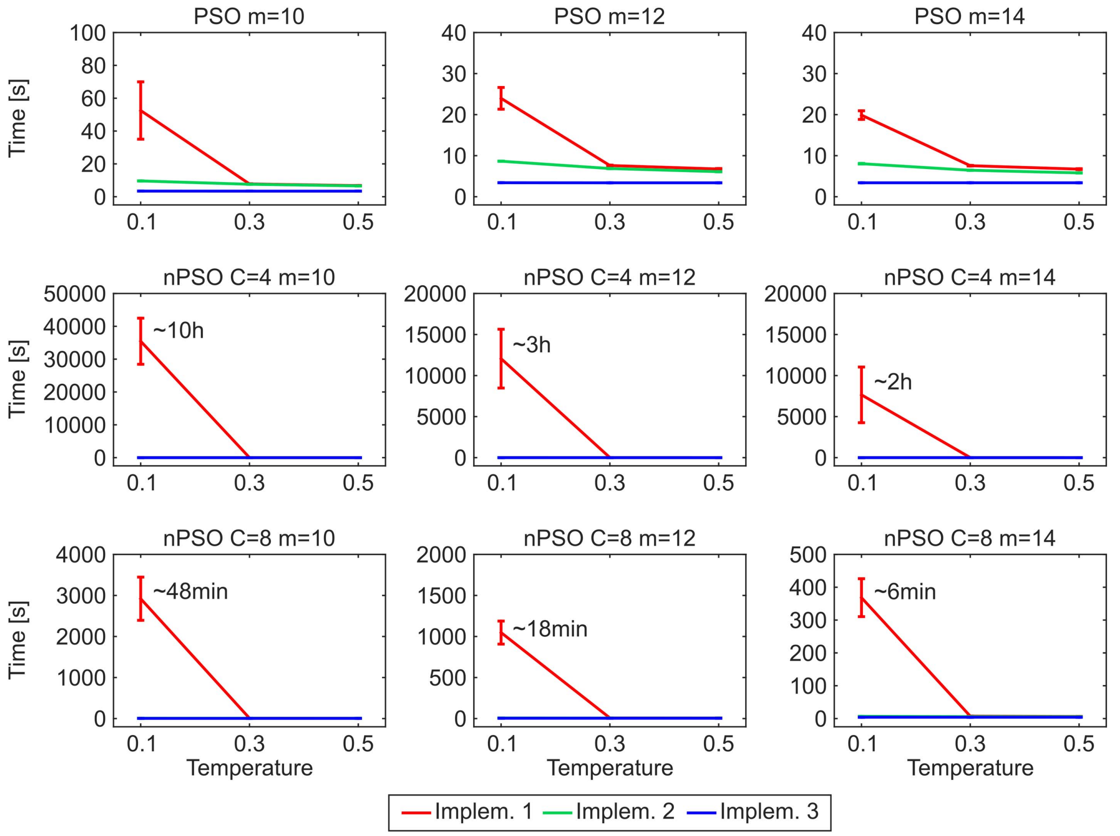


Figure 3. Time performance for generating PSO and nPSO networks. Synthetic networks have been generated using the PSO and nPSO models with parameters $N = 1000$, $m = [10, 12, 14]$, $T = [0.1, 0.3, 0.5]$, $\gamma = 3$ and, for the nPSO model, angular coordinates sampled according to a Gaussian mixture distribution with equal proportions and components $C = [4, 8]$. For each combination of parameters, 10 networks have been generated using the 3 different implementations. The plots report for each parameter combination the mean computational time and standard error over the random repetitions. A more detailed comparison of implementations 2 and 3 and additional plots for different network sizes are provided in supplementary information.

figures 1–3, the computational time for generating PSO networks of size $N = 1000$ is in the order of seconds, whereas for nPSO networks with low temperature $T = 0.1$ it might take almost one hour ($C = 8$) or up to several hours ($C = 4$), depending on the number of communities.

The main reason is the following. Assuming $T > 0$, at each time step i of the generative procedure the new node i picks a randomly chosen existing node $j < i$ and, given that it is not already connected to it, it connects with probability $p(i, j)$, repeating the procedure until it becomes connected to m nodes. However, it is possible to note from equation (1) that the connection probability $p(i, j)$ decreases both for increasing hyperbolic distance and for decreasing temperature (when $h_{ij} > R_i$, which is true in the majority of the cases as shown in supplementary table 1, in particular for increasing network size). Therefore, while generating a network with low temperature and where many nodes are at high hyperbolic distance (for instance: a nPSO model that displays communities presents hyperbolic distances significantly higher than a classical uniform PSO), most of the connection probabilities to the targets will be low. As a consequence, many iterations will be required before that m connections are successfully established. Note that, in the nPSO, the lower the C the higher the distance between adjacent communities, therefore more target nodes will be at high hyperbolic distance, which results in an increased computational time, as pointed out comparing $C = 4$ (supplementary figure 2) and $C = 8$ (supplementary figure 3). Furthermore, in figure 3 it can be noticed that the running time increases also for decreasing m. Although this might result counterintuitive because less links need to be generated, the reason is that for decreasing m the radius R_i of the hyperbolic disk (see equation (1)) decreases, as a consequence also the connection probabilities $p(i, j)$ decrease and therefore more iterations will be required before that m connections are successfully established.

Here we propose two different algorithmic implementations, whose details are provided in the Methods. Figure 3 shows that both the implementations do not present any issue for generating nPSO networks with low temperature. As highlighted in supplementary figures 4–5, the fastest is implementation 3, whose key idea is to sample the target nodes according to the theoretical probabilities $p(i, j)$, and it only requires 5 min to generate large-size nPSO networks of $N = 10\,000$, regardless of the temperature. This is indeed expected since it is the only implementation in which the attempts of establishing new connections are never rejected, and its computational complexity is only dependent on the number of nodes and edges, $O(EN)$. On the contrary, the time complexity of the implementations 1 and 2 has a dependency on the average connection probability to the targets during the generative procedure, which is mainly affected by the temperature and by the extent of the hyperbolic distances. Comparing the complexity of the three implementations and taking into account the computational time of the numerical experiments, we can derive that:

$$O(EN) < O\left(E\left(N + \frac{1}{\tilde{p}_2}\right)\right) < O\left(E\frac{1}{\tilde{p}_1}\right).$$

And therefore:

$$O(N) < O\left(N + \frac{1}{\tilde{p}_2}\right) < O\left(\frac{1}{\tilde{p}_1}\right).$$

This result mainly suggests that, in particular in the scenario of low temperature where the time difference is considerably high, the average number of attempts required by implementation 1 to establish one connection has an order of complexity higher than the number of nodes in the network.

Supplementary figures 6–7 report the time performance for generating GPA networks both with $\Lambda = 0.1$ and $\Lambda = 1$. The advantage of the implementation 3 for low temperature $T = 0.1$ is clearly evident and the computational time difference with implementation 1 becomes more significant for low initial attractiveness $\Lambda = 0.1$. Indeed, in this parameter configuration ($T = 0.1$, $\Lambda = 0.1$) and using implementation 1, the generation of networks with $N = 500$ required several hours and networks with $N = 1000$ were still not generated after one month of running time, whereas implementation 3 required around 1 min for $N = 1000$. We let notice that a lower initial attractiveness tends to locate new coming nodes in regions where other nodes are already present, generating a lower number of denser regions in the hyperbolic disk. The explanation of the higher computational time with respect to $\Lambda = 1$ is therefore analogous to the one given for lower C in the nPSO model.

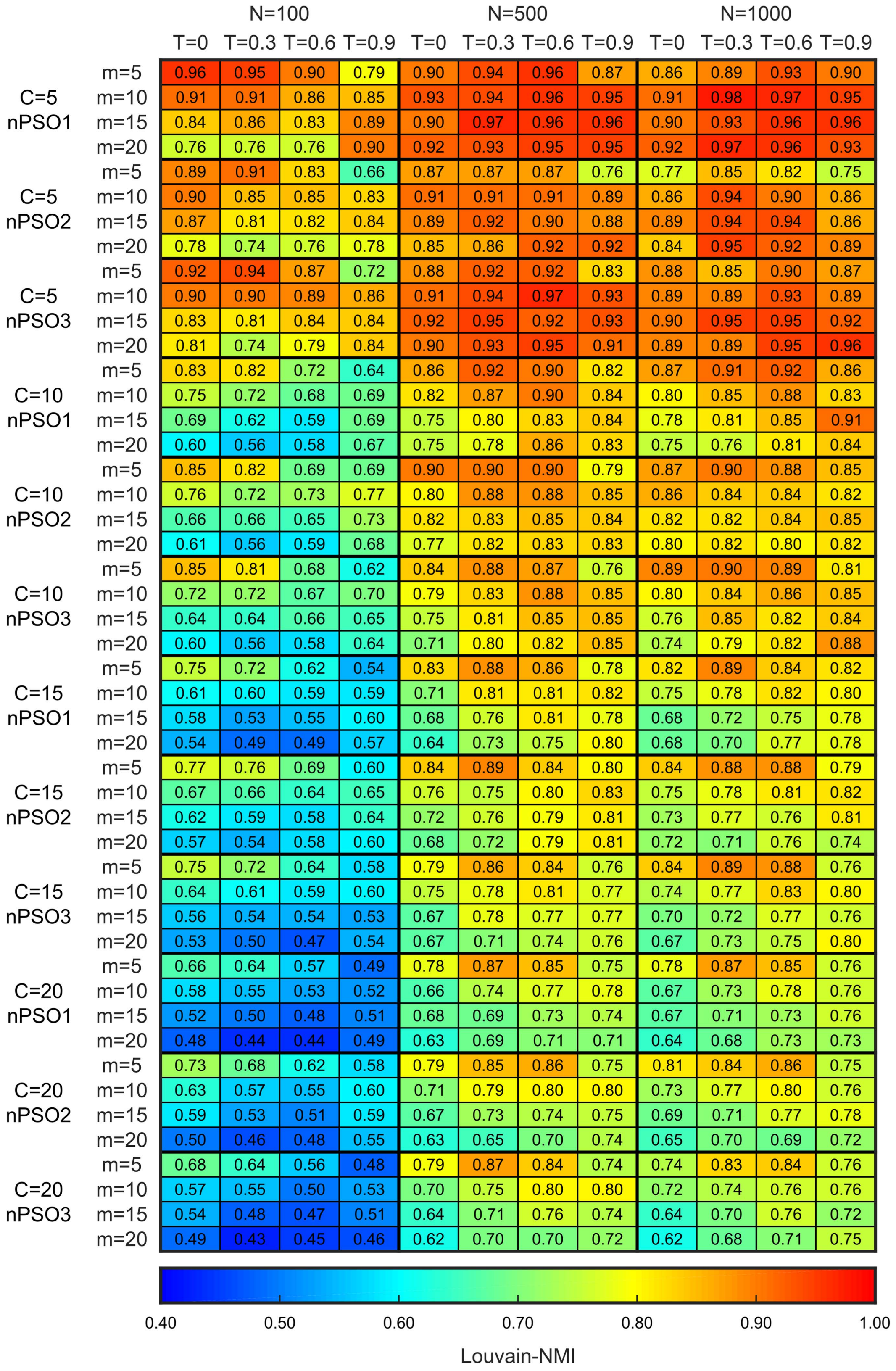
Detectability and mixing property of the nPSO communities

The main novelty introduced by the nPSO model with respect to the GPA model is the possibility to generate a tailored community structure at any given temperature different from $T = 0$. Therefore this section of the paper will lead the reader through a wide investigation on the parameter combinations of the nPSO model for which the emerging communities are detectable by a state-of-the-art algorithm.

Synthetic networks have been generated using the nPSO model with parameters $N = [100, 500, 1000]$, $m = [5, 10, 15, 20]$, $T = [0, 0.3, 0.6, 0.9]$, $\gamma = 3$ and angular coordinates sampled according to mixture distributions with components $C = [5, 10, 15, 20]$ for the three different scenarios previously mentioned: Gaussian mixture with equal proportions (nPSO1), Gaussian mixture with random proportions (nPSO2), Gaussian and Gamma mixture with equal proportions (nPSO3). For more details on the parameters of the mixture distributions please refer to the Methods.

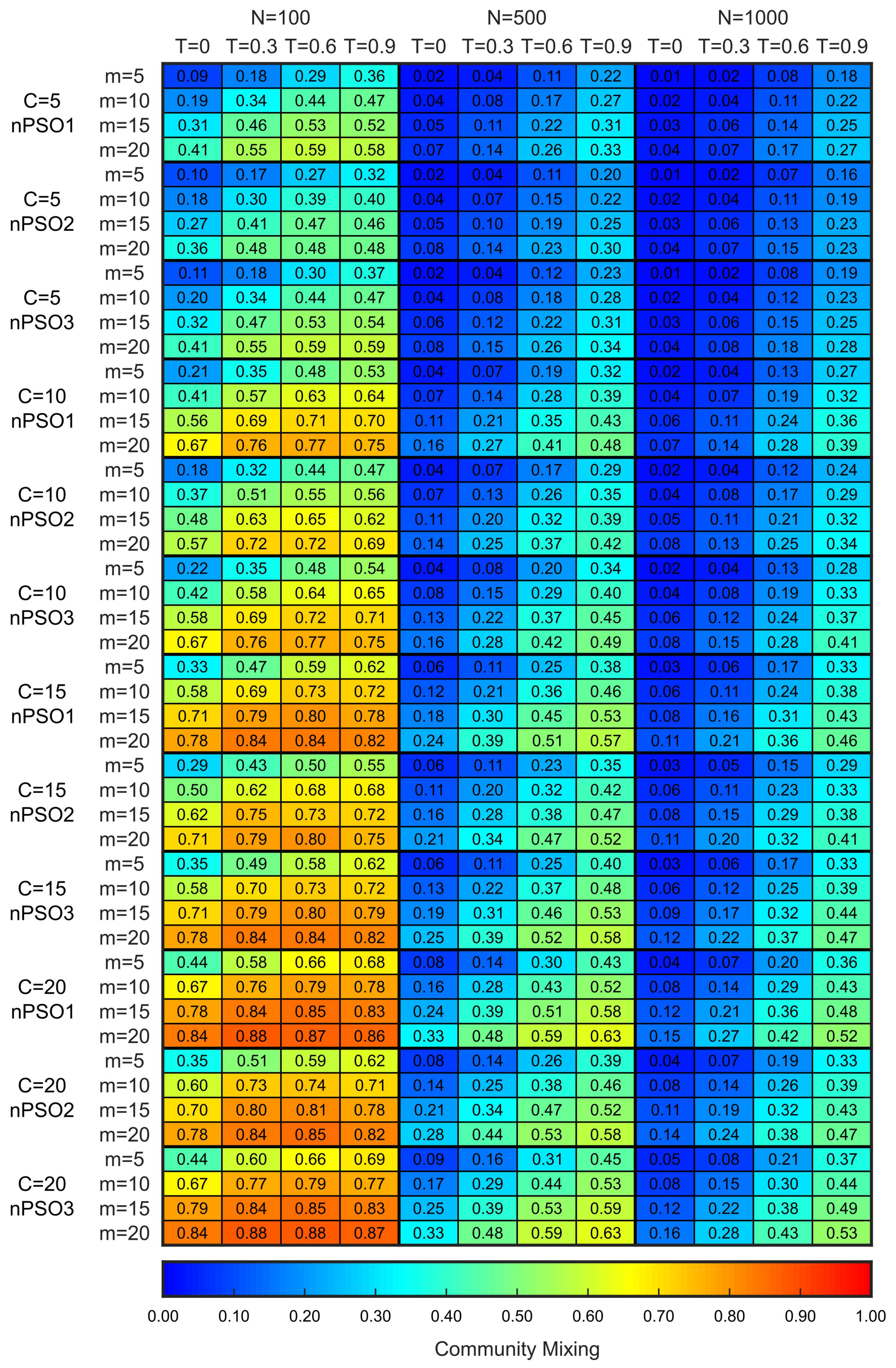
For each network the Louvain community detection algorithm [22] has been executed and the communities detected have been compared to the annotated ones computing the NMI [21] (see supplementary methods for details). We decided to use the Louvain algorithm because it is a model-free and unsupervised heuristic method for community detection based on modularity optimization [22], therefore its performance is not dependent by any assumption on the generative model, and its results should be robust enough regardless of the generative model used to create the synthetic networks. In addition, the Louvain algorithm has been regarded as one of the most effective algorithms for community detection in previous studies across many real and synthetic datasets [23–25]. However, the fact that here we tested the community detectability of the nPSO synthetic networks using only the Louvain algorithm is overcome in a second study where we compare the performance of several state-of-the-art algorithms for community detection across different parameters of the nPSO model [26].

The heatmap in figure 4 reports the mean NMI over 10 repetitions for each parameter combination. The point that firstly captures the attention is the overall higher detectability of the communities in networks of larger size with a lower number of communities (top-right area of the heatmap) in comparison to networks of smaller size with a higher number of communities (bottom-left area of the heatmap). This result suggests that, independently from the kind of mixture distribution (nPSO1, nPSO2 or nPSO3), the ratio between the number of communities (C) and the network size (N) is a factor that strongly affects the detectability of the nPSO communities. This is indeed expected since, for a fixed network size, the lower the number of communities the higher their separation in the angular space. Since connection probabilities depend on geometrical distances, a higher separation leads to a higher percentage of intra-community links with respect to inter-community links (lower community mixing). Previous studies have already demonstrated that the communities are easier to be



		N=100				N=500				N=1000			
		T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9
C=5 nPSO1	m=5	0.96	0.95	0.90	0.79	0.90	0.94	0.96	0.87	0.86	0.89	0.93	0.90
	m=10	0.91	0.91	0.86	0.85	0.93	0.94	0.96	0.95	0.91	0.98	0.97	0.95
	m=15	0.84	0.86	0.83	0.89	0.90	0.97	0.96	0.96	0.90	0.93	0.96	0.96
	m=20	0.76	0.76	0.76	0.90	0.92	0.93	0.95	0.95	0.92	0.97	0.96	0.93
C=5 nPSO2	m=5	0.89	0.91	0.83	0.66	0.87	0.87	0.87	0.76	0.77	0.85	0.82	0.75
	m=10	0.90	0.85	0.85	0.83	0.91	0.91	0.91	0.89	0.86	0.94	0.90	0.86
	m=15	0.87	0.81	0.82	0.84	0.89	0.92	0.90	0.88	0.89	0.94	0.94	0.86
	m=20	0.78	0.74	0.76	0.78	0.85	0.86	0.92	0.92	0.84	0.95	0.92	0.89
C=5 nPSO3	m=5	0.92	0.94	0.87	0.72	0.88	0.92	0.92	0.83	0.88	0.85	0.90	0.87
	m=10	0.90	0.90	0.89	0.86	0.91	0.94	0.97	0.93	0.89	0.89	0.93	0.89
	m=15	0.83	0.81	0.84	0.84	0.92	0.95	0.92	0.93	0.90	0.95	0.95	0.92
	m=20	0.81	0.74	0.79	0.84	0.90	0.93	0.95	0.91	0.89	0.89	0.95	0.96
C=10 nPSO1	m=5	0.83	0.82	0.72	0.64	0.86	0.92	0.90	0.82	0.87	0.91	0.92	0.86
	m=10	0.75	0.72	0.68	0.69	0.82	0.87	0.90	0.84	0.80	0.85	0.88	0.83
	m=15	0.69	0.62	0.59	0.69	0.75	0.80	0.83	0.84	0.78	0.81	0.85	0.91
	m=20	0.60	0.56	0.58	0.67	0.75	0.78	0.86	0.83	0.75	0.76	0.81	0.84
C=10 nPSO2	m=5	0.85	0.82	0.69	0.69	0.90	0.90	0.90	0.79	0.87	0.90	0.88	0.85
	m=10	0.76	0.72	0.73	0.77	0.80	0.88	0.88	0.85	0.86	0.84	0.84	0.82
	m=15	0.66	0.66	0.65	0.73	0.82	0.83	0.85	0.84	0.82	0.82	0.84	0.85
	m=20	0.61	0.56	0.59	0.68	0.77	0.82	0.83	0.83	0.80	0.82	0.80	0.82
C=10 nPSO3	m=5	0.85	0.81	0.68	0.62	0.84	0.88	0.87	0.76	0.89	0.90	0.89	0.81
	m=10	0.72	0.72	0.67	0.70	0.79	0.83	0.88	0.85	0.80	0.84	0.86	0.85
	m=15	0.64	0.64	0.66	0.65	0.75	0.81	0.85	0.85	0.76	0.85	0.82	0.84
	m=20	0.60	0.56	0.58	0.64	0.71	0.80	0.82	0.85	0.74	0.79	0.82	0.88
C=15 nPSO1	m=5	0.75	0.72	0.62	0.54	0.83	0.88	0.86	0.78	0.82	0.89	0.84	0.82
	m=10	0.61	0.60	0.59	0.59	0.71	0.81	0.81	0.82	0.75	0.78	0.82	0.80
	m=15	0.58	0.53	0.55	0.60	0.68	0.76	0.81	0.78	0.68	0.72	0.75	0.78
	m=20	0.54	0.49	0.49	0.57	0.64	0.73	0.75	0.80	0.68	0.70	0.77	0.78
C=15 nPSO2	m=5	0.77	0.76	0.69	0.60	0.84	0.89	0.84	0.80	0.84	0.88	0.88	0.79
	m=10	0.67	0.66	0.64	0.65	0.76	0.75	0.80	0.83	0.75	0.78	0.81	0.82
	m=15	0.62	0.59	0.58	0.64	0.72	0.76	0.79	0.81	0.73	0.77	0.76	0.81
	m=20	0.57	0.54	0.58	0.60	0.68	0.72	0.79	0.81	0.72	0.71	0.76	0.74
C=15 nPSO3	m=5	0.75	0.72	0.64	0.58	0.79	0.86	0.84	0.76	0.84	0.89	0.88	0.76
	m=10	0.64	0.61	0.59	0.60	0.75	0.78	0.81	0.77	0.74	0.77	0.83	0.80
	m=15	0.56	0.54	0.54	0.53	0.67	0.78	0.77	0.77	0.70	0.72	0.77	0.76
	m=20	0.53	0.50	0.47	0.54	0.67	0.71	0.74	0.76	0.67	0.73	0.75	0.80
C=20 nPSO1	m=5	0.66	0.64	0.57	0.49	0.78	0.87	0.85	0.75	0.78	0.87	0.85	0.76
	m=10	0.58	0.55	0.53	0.52	0.66	0.74	0.77	0.78	0.67	0.73	0.78	0.76
	m=15	0.52	0.50	0.48	0.51	0.68	0.69	0.73	0.74	0.67	0.71	0.73	0.76
	m=20	0.48	0.44	0.44	0.49	0.63	0.69	0.71	0.71	0.64	0.68	0.73	0.73
C=20 nPSO2	m=5	0.73	0.68	0.62	0.58	0.79	0.85	0.86	0.75	0.81	0.84	0.86	0.75
	m=10	0.63	0.57	0.55	0.60	0.71	0.79	0.80	0.80	0.73	0.77	0.80	0.76
	m=15	0.59	0.53	0.51	0.59	0.67	0.73	0.74	0.75	0.69	0.71	0.77	0.78
	m=20	0.50	0.46	0.48	0.55	0.63	0.65	0.70	0.74	0.65	0.70	0.69	0.72
C=20 nPSO3	m=5	0.68	0.64	0.56	0.48	0.79	0.87	0.84	0.74	0.74	0.83	0.84	0.76
	m=10	0.57	0.55	0.50	0.53	0.70	0.75	0.80	0.80	0.72	0.74	0.76	0.76
	m=15	0.54	0.48	0.47	0.51	0.64	0.71	0.76	0.74	0.64	0.70	0.76	0.72
	m=20	0.49	0.43	0.45	0.46	0.62	0.70	0.70	0.72	0.62	0.68	0.71	0.75

Figure 4. Detectability of the nPSO communities. Synthetic networks have been generated using the nPSO model with parameters $N = [100, 500, 1000]$, $m = [5, 10, 15, 20]$, $T = [0, 0.3, 0.6, 0.9]$, $\gamma = 3$ and angular coordinates sampled according to mixture distributions of three different kinds with components $C = [5, 10, 15, 20]$: Gaussian mixture with equal proportions (nPSO1), Gaussian mixture with random proportions (nPSO2), Gaussian and Gamma mixture with equal proportions (nPSO3). For more details on the parameters of the mixture distributions please refer to the Methods. For each combination of parameters, 10 networks have been generated. For each network the Louvain community detection method has been executed and the communities detected have been compared to the annotated ones computing the normalized mutual information (NMI). The heatmap reports for each parameter combination the mean NMI, coloured according to a blue-to-red colormap in the range [0.4, 1].



		N=100				N=500				N=1000			
		T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9
C=5 nPSO1	m=5	0.09	0.18	0.29	0.36	0.02	0.04	0.11	0.22	0.01	0.02	0.08	0.18
	m=10	0.19	0.34	0.44	0.47	0.04	0.08	0.17	0.27	0.02	0.04	0.11	0.22
	m=15	0.31	0.46	0.53	0.52	0.05	0.11	0.22	0.31	0.03	0.06	0.14	0.25
	m=20	0.41	0.55	0.59	0.58	0.07	0.14	0.26	0.33	0.04	0.07	0.17	0.27
C=5 nPSO2	m=5	0.10	0.17	0.27	0.32	0.02	0.04	0.11	0.20	0.01	0.02	0.07	0.16
	m=10	0.18	0.30	0.39	0.40	0.04	0.07	0.15	0.22	0.02	0.04	0.11	0.19
	m=15	0.27	0.41	0.47	0.46	0.05	0.10	0.19	0.25	0.03	0.06	0.13	0.23
	m=20	0.36	0.48	0.48	0.48	0.08	0.14	0.23	0.30	0.04	0.07	0.15	0.23
C=5 nPSO3	m=5	0.11	0.18	0.30	0.37	0.02	0.04	0.12	0.23	0.01	0.02	0.08	0.19
	m=10	0.20	0.34	0.44	0.47	0.04	0.08	0.18	0.28	0.02	0.04	0.12	0.23
	m=15	0.32	0.47	0.53	0.54	0.06	0.12	0.22	0.31	0.03	0.06	0.15	0.25
	m=20	0.41	0.55	0.59	0.59	0.08	0.15	0.26	0.34	0.04	0.08	0.18	0.28
C=10 nPSO1	m=5	0.21	0.35	0.48	0.53	0.04	0.07	0.19	0.32	0.02	0.04	0.13	0.27
	m=10	0.41	0.57	0.63	0.64	0.07	0.14	0.28	0.39	0.04	0.07	0.19	0.32
	m=15	0.56	0.69	0.71	0.70	0.11	0.21	0.35	0.43	0.06	0.11	0.24	0.36
	m=20	0.67	0.76	0.77	0.75	0.16	0.27	0.41	0.48	0.07	0.14	0.28	0.39
C=10 nPSO2	m=5	0.18	0.32	0.44	0.47	0.04	0.07	0.17	0.29	0.02	0.04	0.12	0.24
	m=10	0.37	0.51	0.55	0.56	0.07	0.13	0.26	0.35	0.04	0.08	0.17	0.29
	m=15	0.48	0.63	0.65	0.62	0.11	0.20	0.32	0.39	0.05	0.11	0.21	0.32
	m=20	0.57	0.72	0.72	0.69	0.14	0.25	0.37	0.42	0.08	0.13	0.25	0.34
C=10 nPSO3	m=5	0.22	0.35	0.48	0.54	0.04	0.08	0.20	0.34	0.02	0.04	0.13	0.28
	m=10	0.42	0.58	0.64	0.65	0.08	0.15	0.29	0.40	0.04	0.08	0.19	0.33
	m=15	0.58	0.69	0.72	0.71	0.13	0.22	0.37	0.45	0.06	0.12	0.24	0.37
	m=20	0.67	0.76	0.77	0.75	0.16	0.28	0.42	0.49	0.08	0.15	0.28	0.41
C=15 nPSO1	m=5	0.33	0.47	0.59	0.62	0.06	0.11	0.25	0.38	0.03	0.06	0.17	0.33
	m=10	0.58	0.69	0.73	0.72	0.12	0.21	0.36	0.46	0.06	0.11	0.24	0.38
	m=15	0.71	0.79	0.80	0.78	0.18	0.30	0.45	0.53	0.08	0.16	0.31	0.43
	m=20	0.78	0.84	0.84	0.82	0.24	0.39	0.51	0.57	0.11	0.21	0.36	0.46
C=15 nPSO2	m=5	0.29	0.43	0.50	0.55	0.06	0.11	0.23	0.35	0.03	0.05	0.15	0.29
	m=10	0.50	0.62	0.68	0.68	0.11	0.20	0.32	0.42	0.06	0.11	0.23	0.33
	m=15	0.62	0.75	0.73	0.72	0.16	0.28	0.38	0.47	0.08	0.15	0.29	0.38
	m=20	0.71	0.79	0.80	0.75	0.21	0.34	0.47	0.52	0.11	0.20	0.32	0.41
C=15 nPSO3	m=5	0.35	0.49	0.58	0.62	0.06	0.11	0.25	0.40	0.03	0.06	0.17	0.33
	m=10	0.58	0.70	0.73	0.72	0.13	0.22	0.37	0.48	0.06	0.12	0.25	0.39
	m=15	0.71	0.79	0.80	0.79	0.19	0.31	0.46	0.53	0.09	0.17	0.32	0.44
	m=20	0.78	0.84	0.84	0.82	0.25	0.39	0.52	0.58	0.12	0.22	0.37	0.47
C=20 nPSO1	m=5	0.44	0.58	0.66	0.68	0.08	0.14	0.30	0.43	0.04	0.07	0.20	0.36
	m=10	0.67	0.76	0.79	0.78	0.16	0.28	0.43	0.52	0.08	0.14	0.29	0.43
	m=15	0.78	0.84	0.85	0.83	0.24	0.39	0.51	0.58	0.12	0.21	0.36	0.48
	m=20	0.84	0.88	0.87	0.86	0.33	0.48	0.59	0.63	0.15	0.27	0.42	0.52
C=20 nPSO2	m=5	0.35	0.51	0.59	0.62	0.08	0.14	0.26	0.39	0.04	0.07	0.19	0.33
	m=10	0.60	0.73	0.74	0.71	0.14	0.25	0.38	0.46	0.08	0.14	0.26	0.39
	m=15	0.70	0.80	0.81	0.78	0.21	0.34	0.47	0.52	0.11	0.19	0.32	0.43
	m=20	0.78	0.84	0.85	0.82	0.28	0.44	0.53	0.58	0.14	0.24	0.38	0.47
C=20 nPSO3	m=5	0.44	0.60	0.66	0.69	0.09	0.16	0.31	0.45	0.05	0.08	0.21	0.37
	m=10	0.67	0.77	0.79	0.77	0.17	0.29	0.44	0.53	0.08	0.15	0.30	0.44
	m=15	0.79	0.84	0.85	0.83	0.25	0.39	0.53	0.59	0.12	0.22	0.38	0.49
	m=20	0.84	0.88	0.88	0.87	0.33	0.48	0.59	0.63	0.16	0.28	0.43	0.53

Figure 5. Community mixing on nPSO networks. Synthetic networks have been generated using the nPSO model with parameters $N = [100, 500, 1000]$, $m = [5, 10, 15, 20]$, $T = [0, 0.3, 0.6, 0.9]$, $\gamma = 3$ and angular coordinates sampled according to mixture distributions of three different kinds with components $C = [5, 10, 15, 20]$: Gaussian mixture with equal proportions (nPSO1), Gaussian mixture with random proportions (nPSO2), Gaussian and Gamma mixture with equal proportions (nPSO3). For more details on the parameters of the mixture distributions please refer to the Methods. For each combination of parameters, 10 networks have been generated and the community mixing has been computed. The heatmap reports for each parameter combination the mean community mixing, coloured according to a blue-to-red colormap in the range [0, 1].

detected for a lower *community mixing* [24, 27], defined as the average proportion of links from a node to external communities [27].

The heatmap in figure 5 reports the mean community mixing over 10 repetitions for the same parameter combinations as in figure 4. It clearly illustrates that the mixing increases with the number of communities for fixed network size, and it decreases with the network size for fixed number of communities. As expected, the mixing grows also with the temperature, since there is higher probability for a node to establish connections with nodes located far apart from its own community. Furthermore, it increases also with the parameter m, in particular for higher C, where the nodes have too many links with respect to the community size and they are forced to create edges with external communities.

Supplementary figure 8 highlights some particular parameter combinations from the heatmap in figure 4 and this helps to discuss some counterintuitive scenarios due to the parameter combinations. Supplementary figure 8(A) focuses on small networks ($N = 100$) with a low number of communities ($C = 5$) for the first scenario (nPSO1), and it shows that for increasing temperature the NMI decreases for $m = 5$ whereas tends to increase for $m = 20$. This is reasonable because, when each node creates few connections ($m = 5$), directing them towards external communities (higher temperature) makes the community structure less detectable (lower NMI). Instead, when too many links are generated ($m = 20$), a high temperature avoids that most of the inter-communities links are directed to adjacent communities and helps to make more distinct the community boundaries.

Supplementary figure 8(B) reports similar results but from a different perspective. It shows that for increasing m the NMI decreases for $T = 0$ whereas tends to increase for $T = 0.9$. In fact, at $T = 0$ most of the links are internal to the community and increasing m will only increase the links external to the community, being its size small. On the contrary, at $T = 0.9$ many links are also directed to other communities and increasing m will help to have enough internal links to make the communities better detectable. These patterns highlighted on small networks ($N = 100$) with few communities ($C = 5$) are mainly preserved also for higher number of communities, although the overall detectability decreases, as already discussed.

For networks of larger size ($N = [500, 1000]$) with few communities ($C = 5$) the detectability is generally high and tends to be better for middle temperatures. For increasing C the NMI overall decreases, and the highest detectability occurs at low m for middle temperatures (see Supplementary figure 8(C)). The reason is that, due to the larger number of nodes and communities, at $T = 0$ there is higher probability for the nodes to link to adjacent communities, making less distinct the boundaries, while at $T = 0.9$ there is higher probability to lose the preferentially of connection to nodes of the same community. Middle temperatures guarantee a good proportion between links internal to the community and links directed to all the other communities (not only the adjacent ones).

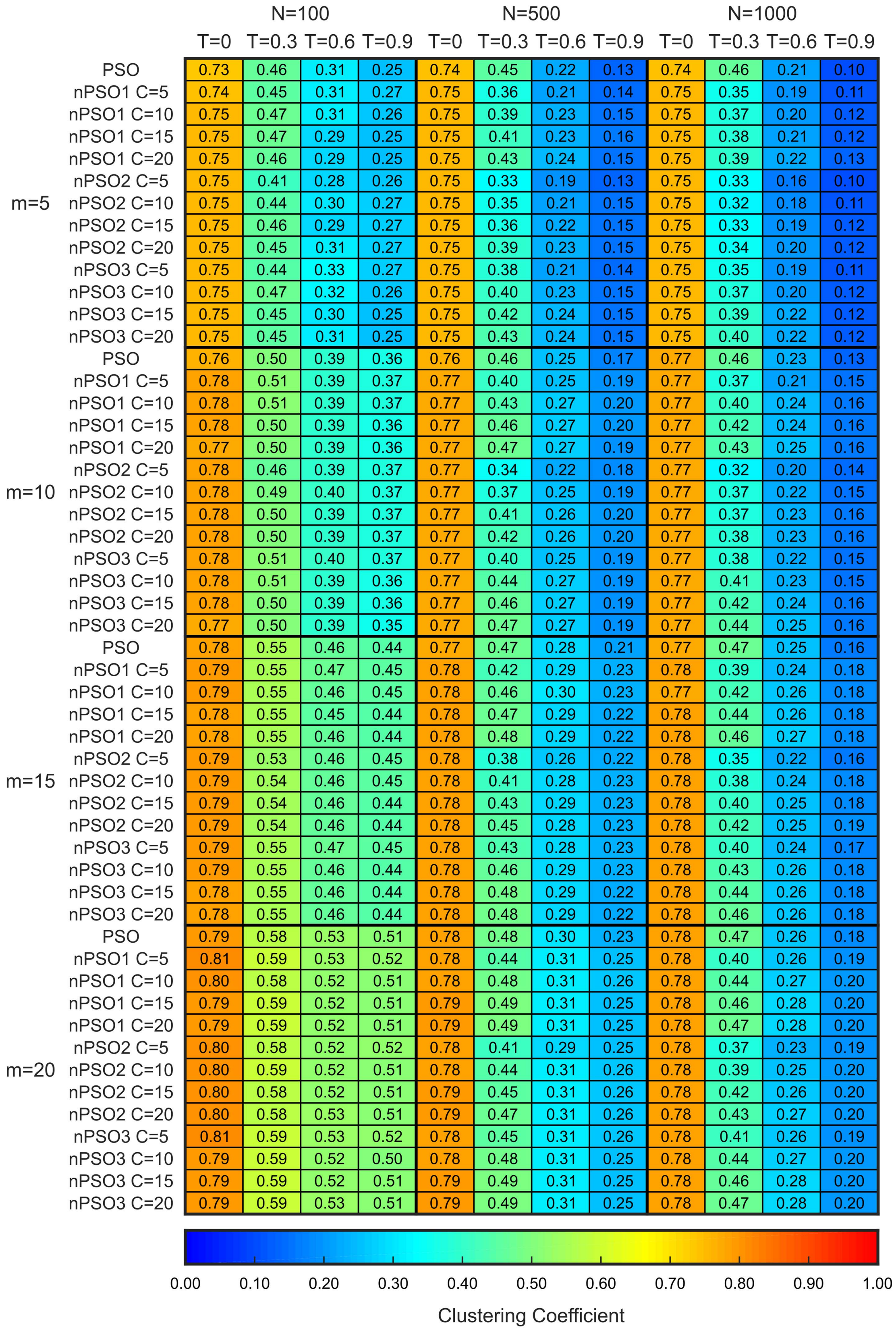
Supplementary figure 8(D) focuses on networks of $N = 1000$ nodes with a high number of communities ($C = 20$) for the first scenario (nPSO1). It shows that for increasing m the NMI decreases for $T = 0$ whereas it is almost not affected for $T = 0.9$. As discussed for the smaller networks, at $T = 0$ most of the links are internal to the community and increasing m will mainly increase the links external to the community, making it less detectable. At $T = 0.9$, differently from what is shown in supplementary figure 8(B), the NMI remains almost constant and does not increase with m, probably due to the fact that the communities are bigger and therefore more internal links are required to make them better detectable.

Finally, although there are some minor variabilities between the different scenarios (nPSO1, nPSO2 and nPSO3), the patterns discussed are mostly consistent over all the nPSO model parameter combinations.

Topological properties of the nPSO networks

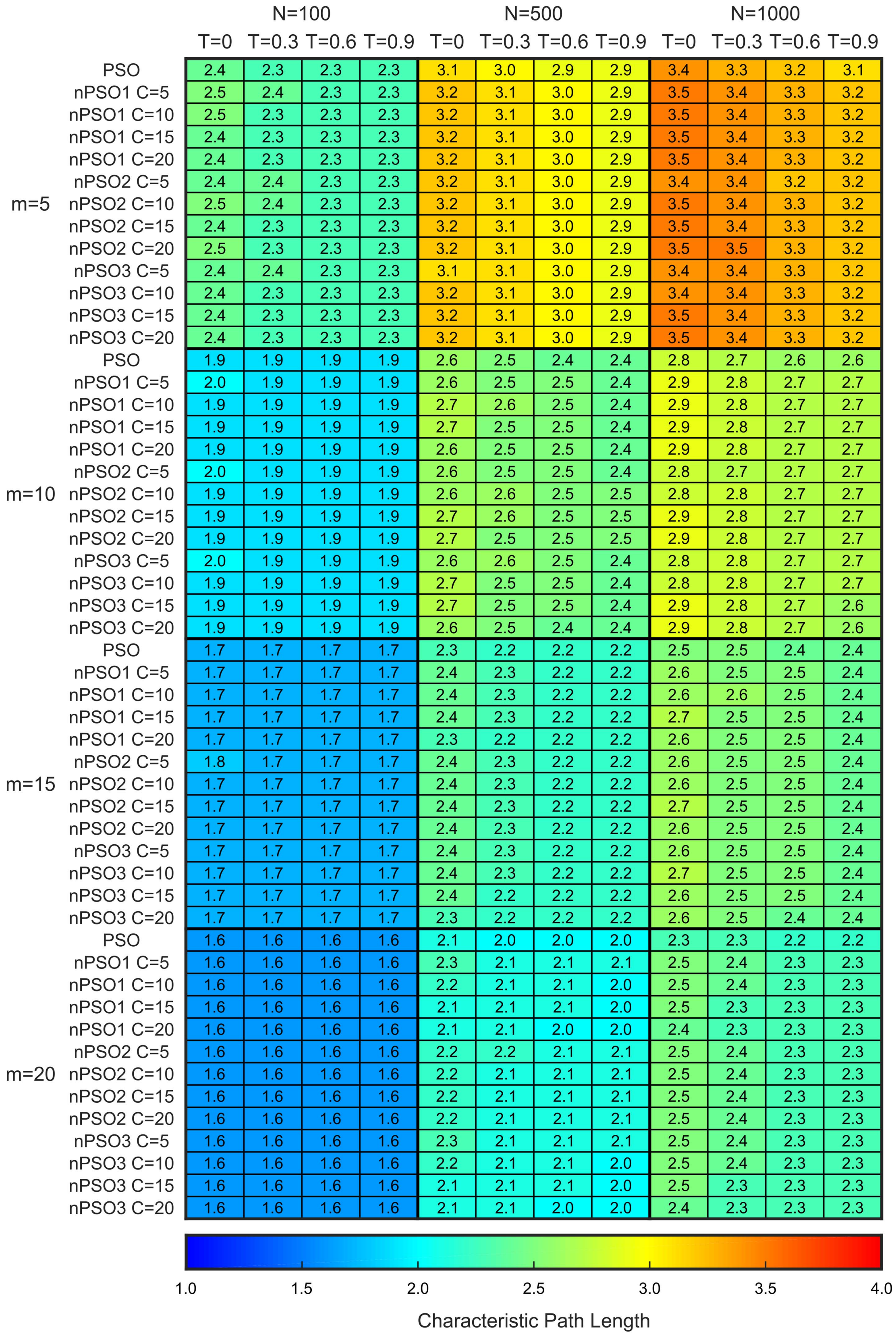
After having proposed a wide investigation on the detectability of the communities generated by the nPSO, in this section we are going to highlight to which extent the community organization affects the main structural properties of the synthetic networks. For the same parameter combinations of the nPSO model as in figure 4, and considering also synthetic networks generated using the (conventional) PSO model with the same parameters N, m, T and γ, we computed several topological measures: clustering coefficient, characteristic path length, assortativity, LCP-correlation, structural consistency, power-law exponent, modularity, small-worldness and rich-clubness. The related heatmaps are reported in figures 6–10 and supplementary figures 9–12 and will be now discussed.

Figure 6 shows the clustering coefficient, which offers an average evaluation of the cross-interaction density between the first neighbours of each node in the network [13]. The clustering coefficient strongly decreases for increasing temperature, since there is higher probability to establish connections between nodes that are far apart from each other, and therefore it is less likely to close triangles in a node neighbourhood. Increasing m tends to increase the clustering coefficient, in fact with a higher number of links it is also more likely to close local triangles, and this is more evident on small networks, where there are less target nodes to connect. The type of



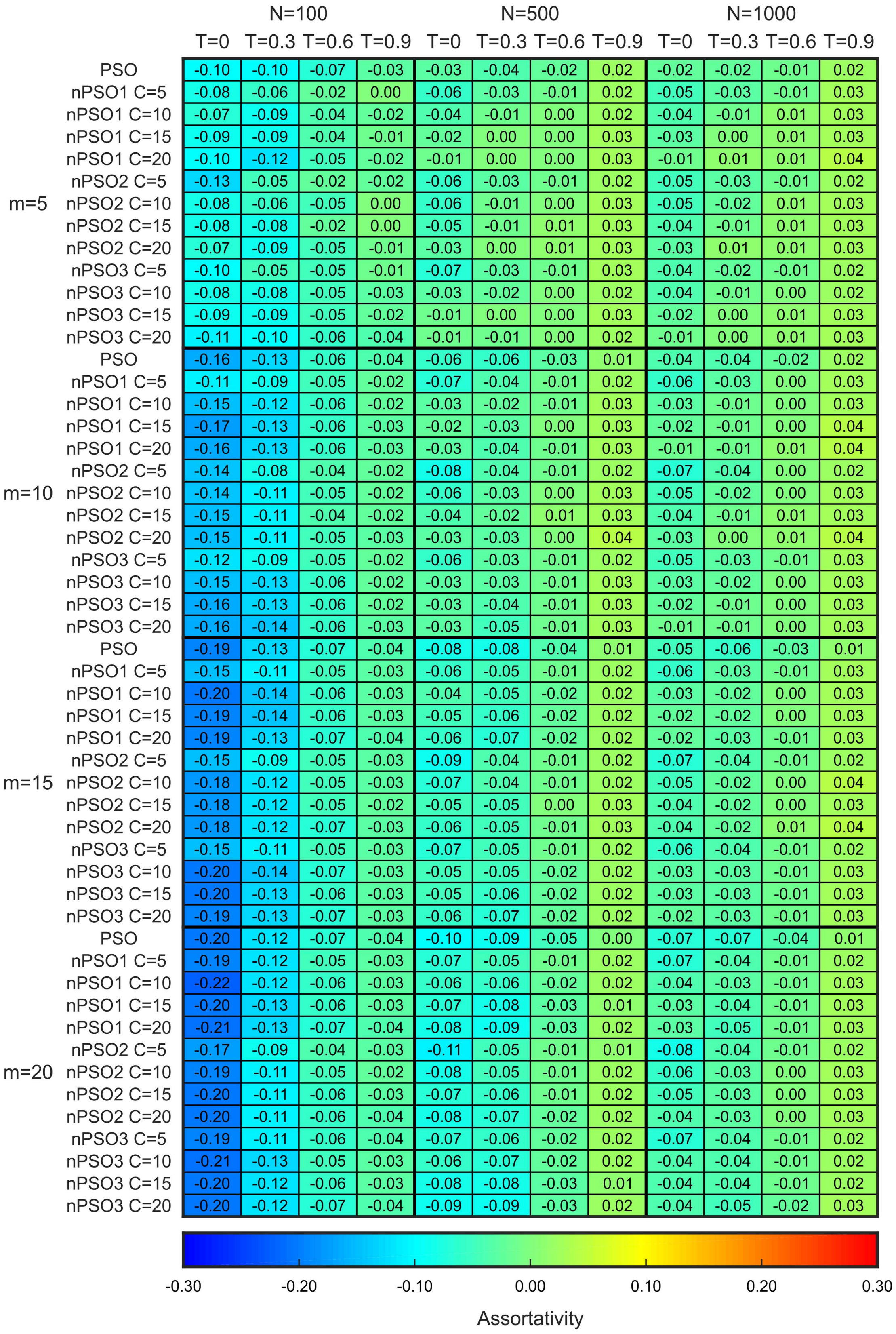
		N=100				N=500				N=1000			
		T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9
m=5	PSO	0.73	0.46	0.31	0.25	0.74	0.45	0.22	0.13	0.74	0.46	0.21	0.10
	nPSO1 C=5	0.74	0.45	0.31	0.27	0.75	0.36	0.21	0.14	0.75	0.35	0.19	0.11
	nPSO1 C=10	0.75	0.47	0.31	0.26	0.75	0.39	0.23	0.15	0.75	0.37	0.20	0.12
	nPSO1 C=15	0.75	0.47	0.29	0.25	0.75	0.41	0.23	0.16	0.75	0.38	0.21	0.12
	nPSO1 C=20	0.75	0.46	0.29	0.25	0.75	0.43	0.24	0.15	0.75	0.39	0.22	0.13
	nPSO2 C=5	0.75	0.41	0.28	0.26	0.75	0.33	0.19	0.13	0.75	0.33	0.16	0.10
	nPSO2 C=10	0.75	0.44	0.30	0.27	0.75	0.35	0.21	0.15	0.75	0.32	0.18	0.11
	nPSO2 C=15	0.75	0.46	0.29	0.27	0.75	0.36	0.22	0.15	0.75	0.33	0.19	0.12
	nPSO2 C=20	0.75	0.45	0.31	0.27	0.75	0.39	0.23	0.15	0.75	0.34	0.20	0.12
	nPSO3 C=5	0.75	0.44	0.33	0.27	0.75	0.38	0.21	0.14	0.75	0.35	0.19	0.11
	nPSO3 C=10	0.75	0.47	0.32	0.26	0.75	0.40	0.23	0.15	0.75	0.37	0.20	0.12
	nPSO3 C=15	0.75	0.45	0.30	0.25	0.75	0.42	0.24	0.15	0.75	0.39	0.22	0.12
	nPSO3 C=20	0.75	0.45	0.31	0.25	0.75	0.43	0.24	0.15	0.75	0.40	0.22	0.12
m=10	PSO	0.76	0.50	0.39	0.36	0.76	0.46	0.25	0.17	0.77	0.46	0.23	0.13
	nPSO1 C=5	0.78	0.51	0.39	0.37	0.77	0.40	0.25	0.19	0.77	0.37	0.21	0.15
	nPSO1 C=10	0.78	0.51	0.39	0.37	0.77	0.43	0.27	0.20	0.77	0.40	0.24	0.16
	nPSO1 C=15	0.78	0.50	0.39	0.36	0.77	0.46	0.27	0.20	0.77	0.42	0.24	0.16
	nPSO1 C=20	0.77	0.50	0.39	0.36	0.77	0.47	0.27	0.19	0.77	0.43	0.25	0.16
	nPSO2 C=5	0.78	0.46	0.39	0.37	0.77	0.34	0.22	0.18	0.77	0.32	0.20	0.14
	nPSO2 C=10	0.78	0.49	0.40	0.37	0.77	0.37	0.25	0.19	0.77	0.37	0.22	0.15
	nPSO2 C=15	0.78	0.50	0.39	0.37	0.77	0.41	0.26	0.20	0.77	0.37	0.23	0.16
	nPSO2 C=20	0.78	0.50	0.39	0.37	0.77	0.42	0.26	0.20	0.77	0.38	0.23	0.16
	nPSO3 C=5	0.78	0.51	0.40	0.37	0.77	0.40	0.25	0.19	0.77	0.38	0.22	0.15
	nPSO3 C=10	0.78	0.51	0.39	0.36	0.77	0.44	0.27	0.19	0.77	0.41	0.23	0.15
	nPSO3 C=15	0.78	0.50	0.39	0.36	0.77	0.46	0.27	0.19	0.77	0.42	0.24	0.16
	nPSO3 C=20	0.77	0.50	0.39	0.35	0.77	0.47	0.27	0.19	0.77	0.44	0.25	0.16
m=15	PSO	0.78	0.55	0.46	0.44	0.77	0.47	0.28	0.21	0.77	0.47	0.25	0.16
	nPSO1 C=5	0.79	0.55	0.47	0.45	0.78	0.42	0.29	0.23	0.78	0.39	0.24	0.18
	nPSO1 C=10	0.79	0.55	0.46	0.45	0.78	0.46	0.30	0.23	0.77	0.42	0.26	0.18
	nPSO1 C=15	0.78	0.55	0.45	0.44	0.78	0.47	0.29	0.22	0.78	0.44	0.26	0.18
	nPSO1 C=20	0.78	0.55	0.46	0.44	0.78	0.48	0.29	0.22	0.78	0.46	0.27	0.18
	nPSO2 C=5	0.79	0.53	0.46	0.45	0.78	0.38	0.26	0.22	0.78	0.35	0.22	0.16
	nPSO2 C=10	0.79	0.54	0.46	0.45	0.78	0.41	0.28	0.23	0.78	0.38	0.24	0.18
	nPSO2 C=15	0.79	0.54	0.46	0.44	0.78	0.43	0.29	0.23	0.78	0.40	0.25	0.18
	nPSO2 C=20	0.79	0.54	0.46	0.44	0.78	0.45	0.28	0.23	0.78	0.42	0.25	0.19
	nPSO3 C=5	0.79	0.55	0.47	0.45	0.78	0.43	0.28	0.23	0.78	0.40	0.24	0.17
	nPSO3 C=10	0.79	0.55	0.46	0.44	0.78	0.46	0.29	0.23	0.78	0.43	0.26	0.18
	nPSO3 C=15	0.78	0.55	0.46	0.44	0.78	0.48	0.29	0.22	0.78	0.44	0.26	0.18
	nPSO3 C=20	0.78	0.55	0.46	0.44	0.78	0.48	0.29	0.22	0.78	0.46	0.26	0.18
m=20	PSO	0.79	0.58	0.53	0.51	0.78	0.48	0.30	0.23	0.78	0.47	0.26	0.18
	nPSO1 C=5	0.81	0.59	0.53	0.52	0.78	0.44	0.31	0.25	0.78	0.40	0.26	0.19
	nPSO1 C=10	0.80	0.58	0.52	0.51	0.78	0.48	0.31	0.26	0.78	0.44	0.27	0.20
	nPSO1 C=15	0.79	0.59	0.52	0.51	0.79	0.49	0.31	0.25	0.78	0.46	0.28	0.20
	nPSO1 C=20	0.79	0.59	0.52	0.51	0.79	0.49	0.31	0.25	0.78	0.47	0.28	0.20
	nPSO2 C=5	0.80	0.58	0.52	0.52	0.78	0.41	0.29	0.25	0.78	0.37	0.23	0.19
	nPSO2 C=10	0.80	0.59	0.52	0.51	0.78	0.44	0.31	0.26	0.78	0.39	0.25	0.20
	nPSO2 C=15	0.80	0.58	0.52	0.51	0.79	0.45	0.31	0.26	0.78	0.42	0.26	0.20
	nPSO2 C=20	0.80	0.58	0.53	0.51	0.79	0.47	0.31	0.26	0.78	0.43	0.27	0.20
	nPSO3 C=5	0.81	0.59	0.53	0.52	0.78	0.45	0.31	0.26	0.78	0.41	0.26	0.19
	nPSO3 C=10	0.79	0.59	0.52	0.50	0.78	0.48	0.31	0.25	0.78	0.44	0.27	0.20
	nPSO3 C=15	0.79	0.59	0.52	0.51	0.79	0.49	0.31	0.25	0.78	0.46	0.28	0.20
	nPSO3 C=20	0.79	0.59	0.53	0.51	0.79	0.49	0.31	0.25	0.78	0.47	0.28	0.20

Figure 6. Clustering coefficient of the PSO and nPSO networks. Synthetic networks have been generated using the nPSO model with parameters $N = [100, 500, 1000]$, $m = [5, 10, 15, 20]$, $T = [0, 0.3, 0.6, 0.9]$, $\gamma = 3$ and angular coordinates sampled according to mixture distributions of three different kinds with components $C = [5, 10, 15, 20]$: Gaussian mixture with equal proportions (nPSO1), Gaussian mixture with random proportions (nPSO2), Gaussian and Gamma mixture with equal proportions (nPSO3). For more details on the parameters of the mixture distributions please refer to the Methods. Furthermore, synthetic networks have been generated also using the PSO model with the same parameters N, m, T and γ. For each combination of parameters, 10 networks have been generated and the clustering coefficient has been computed. The heatmap reports for each parameter combination the mean clustering coefficient, coloured according to a blue-to-red colormap in the range [0, 1].



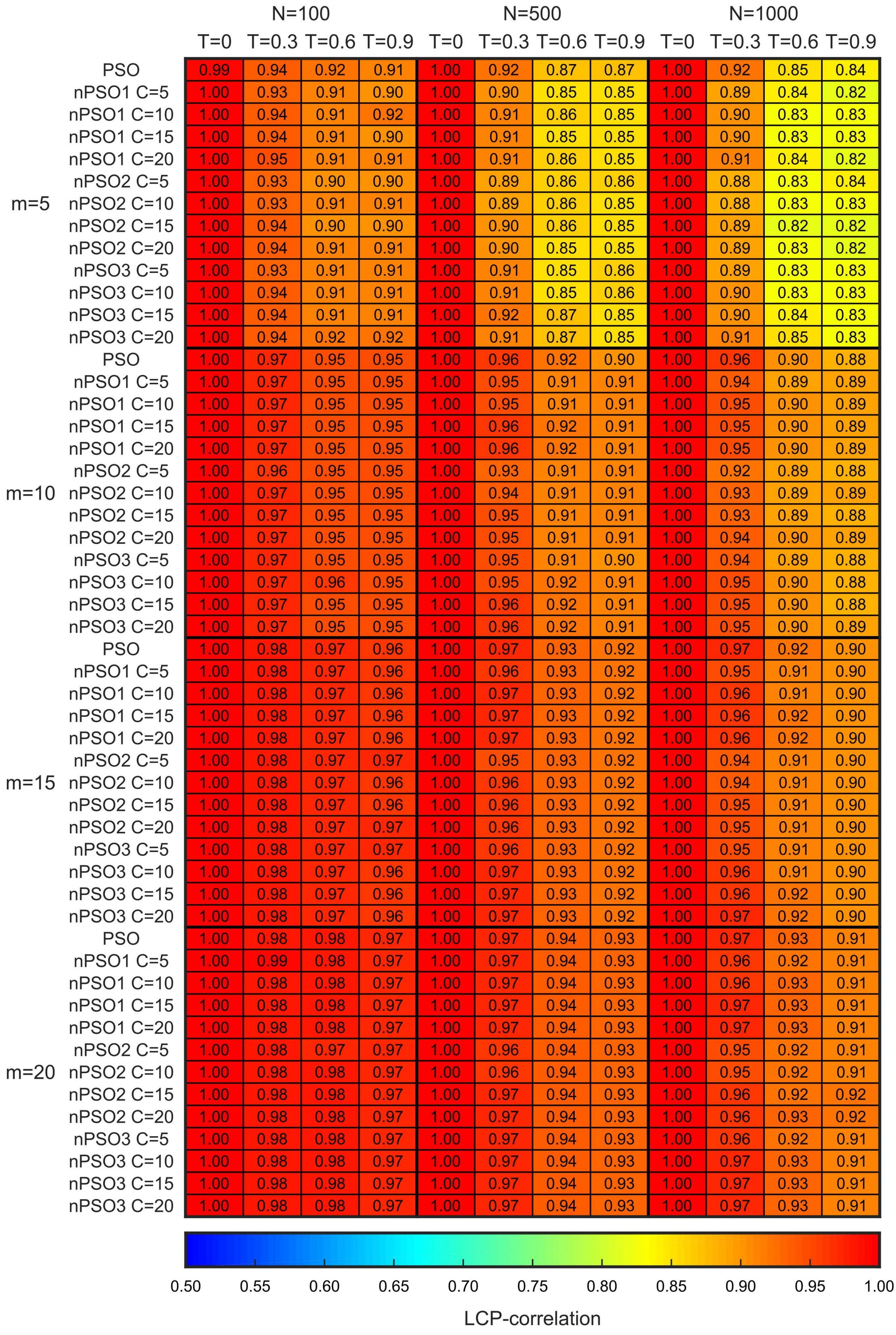
		N=100				N=500				N=1000			
		T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9
m=5	PSO	2.4	2.3	2.3	2.3	3.1	3.0	2.9	2.9	3.4	3.3	3.2	3.1
	nPSO1 C=5	2.5	2.4	2.3	2.3	3.2	3.1	3.0	2.9	3.5	3.4	3.3	3.2
	nPSO1 C=10	2.5	2.3	2.3	2.3	3.2	3.1	3.0	2.9	3.5	3.4	3.3	3.2
	nPSO1 C=15	2.4	2.3	2.3	2.3	3.2	3.1	3.0	2.9	3.5	3.4	3.3	3.2
	nPSO1 C=20	2.4	2.3	2.3	2.3	3.2	3.1	3.0	2.9	3.5	3.4	3.3	3.2
	nPSO2 C=5	2.4	2.4	2.3	2.3	3.2	3.1	3.0	2.9	3.4	3.4	3.2	3.2
	nPSO2 C=10	2.5	2.4	2.3	2.3	3.2	3.1	3.0	2.9	3.5	3.4	3.3	3.2
	nPSO2 C=15	2.4	2.3	2.3	2.3	3.2	3.1	3.0	2.9	3.5	3.4	3.3	3.2
	nPSO2 C=20	2.5	2.3	2.3	2.3	3.2	3.1	3.0	2.9	3.5	3.5	3.3	3.2
	nPSO3 C=5	2.4	2.4	2.3	2.3	3.1	3.1	3.0	2.9	3.4	3.4	3.3	3.2
	nPSO3 C=10	2.4	2.3	2.3	2.3	3.2	3.1	3.0	2.9	3.4	3.4	3.3	3.2
	nPSO3 C=15	2.4	2.3	2.3	2.3	3.2	3.1	3.0	2.9	3.5	3.4	3.3	3.2
	nPSO3 C=20	2.4	2.3	2.3	2.3	3.2	3.1	3.0	2.9	3.5	3.4	3.3	3.2
m=10	PSO	1.9	1.9	1.9	1.9	2.6	2.5	2.4	2.4	2.8	2.7	2.6	2.6
	nPSO1 C=5	2.0	1.9	1.9	1.9	2.6	2.5	2.5	2.4	2.9	2.8	2.7	2.7
	nPSO1 C=10	1.9	1.9	1.9	1.9	2.7	2.6	2.5	2.4	2.9	2.8	2.7	2.7
	nPSO1 C=15	1.9	1.9	1.9	1.9	2.7	2.5	2.5	2.4	2.9	2.8	2.7	2.7
	nPSO1 C=20	1.9	1.9	1.9	1.9	2.6	2.5	2.5	2.4	2.9	2.8	2.7	2.7
	nPSO2 C=5	2.0	1.9	1.9	1.9	2.6	2.5	2.5	2.4	2.8	2.7	2.7	2.7
	nPSO2 C=10	1.9	1.9	1.9	1.9	2.6	2.6	2.5	2.5	2.8	2.8	2.7	2.7
	nPSO2 C=15	1.9	1.9	1.9	1.9	2.7	2.6	2.5	2.5	2.9	2.8	2.7	2.7
	nPSO2 C=20	1.9	1.9	1.9	1.9	2.7	2.5	2.5	2.5	2.9	2.8	2.7	2.7
	nPSO3 C=5	2.0	1.9	1.9	1.9	2.6	2.6	2.5	2.4	2.8	2.8	2.7	2.7
	nPSO3 C=10	1.9	1.9	1.9	1.9	2.7	2.5	2.5	2.4	2.8	2.8	2.7	2.7
	nPSO3 C=15	1.9	1.9	1.9	1.9	2.7	2.5	2.5	2.4	2.9	2.8	2.7	2.6
	nPSO3 C=20	1.9	1.9	1.9	1.9	2.6	2.5	2.4	2.4	2.9	2.8	2.7	2.6
m=15	PSO	1.7	1.7	1.7	1.7	2.3	2.2	2.2	2.2	2.5	2.5	2.4	2.4
	nPSO1 C=5	1.7	1.7	1.7	1.7	2.4	2.3	2.2	2.2	2.6	2.5	2.5	2.4
	nPSO1 C=10	1.7	1.7	1.7	1.7	2.4	2.3	2.2	2.2	2.6	2.6	2.5	2.4
	nPSO1 C=15	1.7	1.7	1.7	1.7	2.4	2.3	2.2	2.2	2.7	2.5	2.5	2.4
	nPSO1 C=20	1.7	1.7	1.7	1.7	2.3	2.2	2.2	2.2	2.6	2.5	2.5	2.4
	nPSO2 C=5	1.8	1.7	1.7	1.7	2.4	2.3	2.2	2.2	2.6	2.5	2.5	2.4
	nPSO2 C=10	1.7	1.7	1.7	1.7	2.4	2.3	2.2	2.2	2.6	2.5	2.5	2.4
	nPSO2 C=15	1.7	1.7	1.7	1.7	2.4	2.3	2.2	2.2	2.7	2.5	2.5	2.4
	nPSO2 C=20	1.7	1.7	1.7	1.7	2.4	2.3	2.2	2.2	2.6	2.5	2.5	2.4
	nPSO3 C=5	1.7	1.7	1.7	1.7	2.4	2.3	2.2	2.2	2.6	2.5	2.5	2.4
	nPSO3 C=10	1.7	1.7	1.7	1.7	2.4	2.3	2.2	2.2	2.7	2.5	2.5	2.4
	nPSO3 C=15	1.7	1.7	1.7	1.7	2.4	2.2	2.2	2.2	2.6	2.5	2.5	2.4
	nPSO3 C=20	1.7	1.7	1.7	1.7	2.3	2.2	2.2	2.2	2.6	2.5	2.4	2.4
m=20	PSO	1.6	1.6	1.6	1.6	2.1	2.0	2.0	2.0	2.3	2.3	2.2	2.2
	nPSO1 C=5	1.6	1.6	1.6	1.6	2.3	2.1	2.1	2.1	2.5	2.4	2.3	2.3
	nPSO1 C=10	1.6	1.6	1.6	1.6	2.2	2.1	2.1	2.0	2.5	2.4	2.3	2.3
	nPSO1 C=15	1.6	1.6	1.6	1.6	2.1	2.1	2.1	2.0	2.5	2.3	2.3	2.3
	nPSO1 C=20	1.6	1.6	1.6	1.6	2.1	2.1	2.0	2.0	2.4	2.3	2.3	2.3
	nPSO2 C=5	1.6	1.6	1.6	1.6	2.2	2.2	2.1	2.1	2.5	2.4	2.3	2.3
	nPSO2 C=10	1.6	1.6	1.6	1.6	2.2	2.1	2.1	2.1	2.5	2.4	2.3	2.3
	nPSO2 C=15	1.6	1.6	1.6	1.6	2.2	2.1	2.1	2.1	2.5	2.4	2.3	2.3
	nPSO2 C=20	1.6	1.6	1.6	1.6	2.2	2.1	2.1	2.1	2.5	2.4	2.3	2.3
	nPSO3 C=5	1.6	1.6	1.6	1.6	2.3	2.1	2.1	2.1	2.5	2.4	2.3	2.3
	nPSO3 C=10	1.6	1.6	1.6	1.6	2.2	2.1	2.1	2.0	2.5	2.4	2.3	2.3
	nPSO3 C=15	1.6	1.6	1.6	1.6	2.1	2.1	2.1	2.0	2.5	2.3	2.3	2.3
	nPSO3 C=20	1.6	1.6	1.6	1.6	2.1	2.1	2.0	2.0	2.4	2.3	2.3	2.3

Figure 7. Characteristic path length of the PSO and nPSO networks. Synthetic networks have been generated using the nPSO model with parameters $N = [100, 500, 1000]$, $m = [5, 10, 15, 20]$, $T = [0, 0.3, 0.6, 0.9]$, $\gamma = 3$ and angular coordinates sampled according to mixture distributions of three different kinds with components $C = [5, 10, 15, 20]$: Gaussian mixture with equal proportions (nPSO1), Gaussian mixture with random proportions (nPSO2), Gaussian and Gamma mixture with equal proportions (nPSO3). For more details on the parameters of the mixture distributions please refer to the Methods. Furthermore, synthetic networks have been generated also using the PSO model with the same parameters N, m, T and γ. For each combination of parameters, 10 networks have been generated and the characteristic path length has been computed. The heatmap reports for each parameter combination the mean characteristic path length, coloured according to a blue-to-red colormap in the range [1, 4].



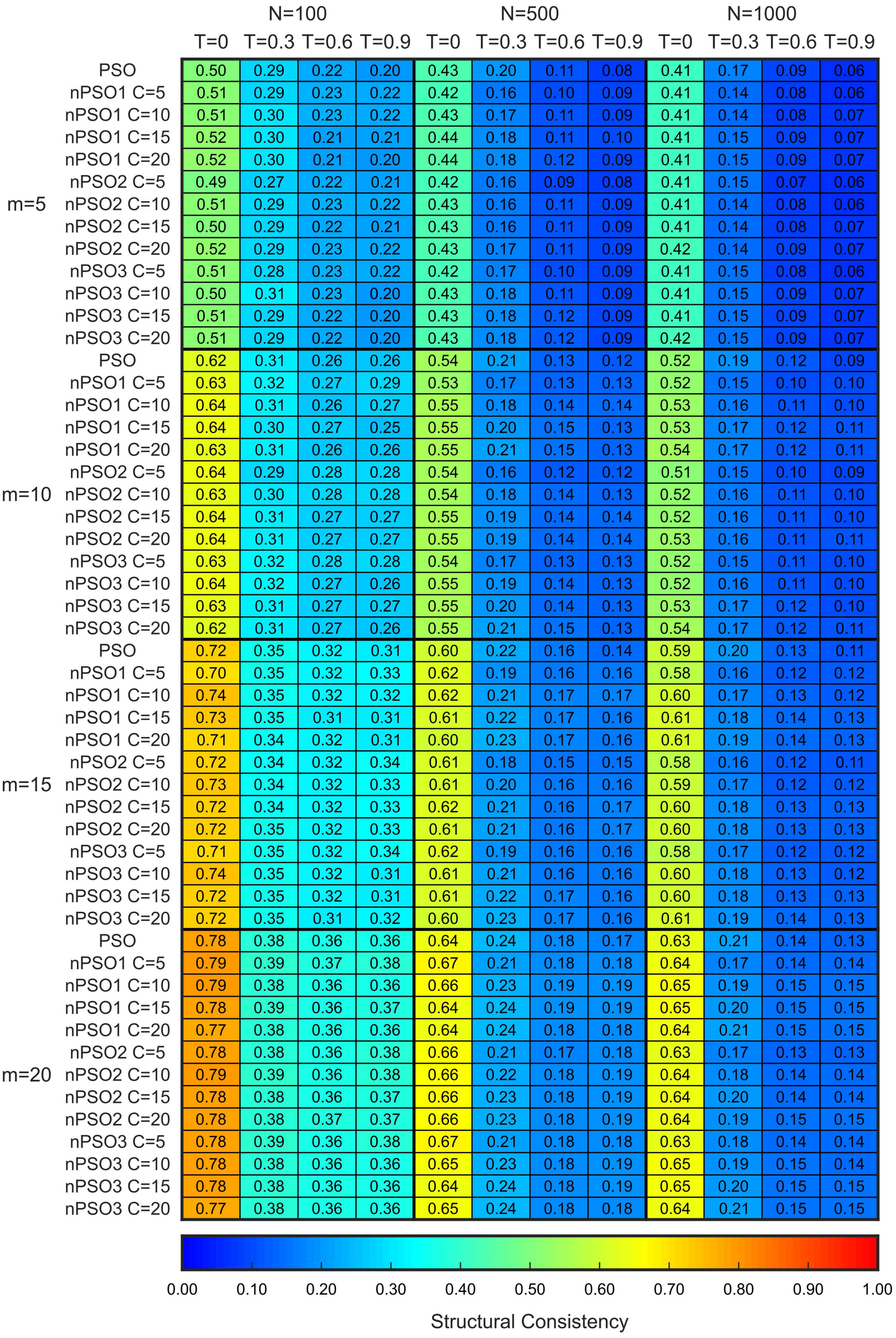
		N=100				N=500				N=1000			
		T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9
m=5	PSO	-0.10	-0.10	-0.07	-0.03	-0.03	-0.04	-0.02	0.02	-0.02	-0.02	-0.01	0.02
	nPSO1 C=5	-0.08	-0.06	-0.02	0.00	-0.06	-0.03	-0.01	0.02	-0.05	-0.03	-0.01	0.03
	nPSO1 C=10	-0.07	-0.09	-0.04	-0.02	-0.04	-0.01	0.00	0.02	-0.04	-0.01	0.01	0.03
	nPSO1 C=15	-0.09	-0.09	-0.04	-0.01	-0.02	0.00	0.00	0.03	-0.03	0.00	0.01	0.03
	nPSO1 C=20	-0.10	-0.12	-0.05	-0.02	-0.01	0.00	0.00	0.03	-0.01	0.01	0.01	0.04
	nPSO2 C=5	-0.13	-0.05	-0.02	-0.02	-0.06	-0.03	-0.01	0.02	-0.05	-0.03	-0.01	0.02
	nPSO2 C=10	-0.08	-0.06	-0.05	0.00	-0.06	-0.01	0.00	0.03	-0.05	-0.02	0.01	0.03
	nPSO2 C=15	-0.08	-0.08	-0.02	0.00	-0.05	-0.01	0.01	0.03	-0.04	-0.01	0.01	0.03
	nPSO2 C=20	-0.07	-0.09	-0.05	-0.01	-0.03	0.00	0.01	0.03	-0.03	0.01	0.01	0.03
	nPSO3 C=5	-0.10	-0.05	-0.05	-0.01	-0.07	-0.03	-0.01	0.03	-0.04	-0.02	-0.01	0.02
	nPSO3 C=10	-0.08	-0.08	-0.05	-0.03	-0.03	-0.02	0.00	0.02	-0.04	-0.01	0.00	0.02
	nPSO3 C=15	-0.09	-0.09	-0.05	-0.02	-0.01	0.00	0.00	0.03	-0.02	0.00	0.01	0.03
	nPSO3 C=20	-0.11	-0.10	-0.06	-0.04	-0.01	-0.01	0.00	0.02	-0.01	0.00	0.01	0.03
m=10	PSO	-0.16	-0.13	-0.06	-0.04	-0.06	-0.06	-0.03	0.01	-0.04	-0.04	-0.02	0.02
	nPSO1 C=5	-0.11	-0.09	-0.05	-0.02	-0.07	-0.04	-0.01	0.02	-0.06	-0.03	0.00	0.03
	nPSO1 C=10	-0.15	-0.12	-0.06	-0.02	-0.03	-0.02	-0.01	0.03	-0.03	-0.01	0.00	0.03
	nPSO1 C=15	-0.17	-0.13	-0.06	-0.03	-0.02	-0.03	0.00	0.03	-0.02	-0.01	0.00	0.04
	nPSO1 C=20	-0.16	-0.13	-0.06	-0.03	-0.03	-0.04	-0.01	0.03	-0.01	-0.01	0.01	0.04
	nPSO2 C=5	-0.14	-0.08	-0.04	-0.02	-0.08	-0.04	-0.01	0.02	-0.07	-0.04	0.00	0.02
	nPSO2 C=10	-0.14	-0.11	-0.05	-0.02	-0.06	-0.03	0.00	0.03	-0.05	-0.02	0.00	0.03
	nPSO2 C=15	-0.15	-0.11	-0.04	-0.02	-0.04	-0.02	0.01	0.03	-0.04	-0.01	0.01	0.03
	nPSO2 C=20	-0.15	-0.11	-0.05	-0.03	-0.03	-0.03	0.00	0.04	-0.03	0.00	0.01	0.04
	nPSO3 C=5	-0.12	-0.09	-0.05	-0.02	-0.06	-0.03	-0.01	0.02	-0.05	-0.03	-0.01	0.03
	nPSO3 C=10	-0.15	-0.13	-0.06	-0.02	-0.03	-0.03	-0.01	0.03	-0.03	-0.02	0.00	0.03
	nPSO3 C=15	-0.16	-0.13	-0.06	-0.02	-0.03	-0.04	-0.01	0.03	-0.02	-0.01	0.00	0.03
	nPSO3 C=20	-0.16	-0.14	-0.06	-0.03	-0.03	-0.05	-0.01	0.03	-0.01	-0.01	0.00	0.03
m=15	PSO	-0.19	-0.13	-0.07	-0.04	-0.08	-0.08	-0.04	0.01	-0.05	-0.06	-0.03	0.01
	nPSO1 C=5	-0.15	-0.11	-0.05	-0.03	-0.06	-0.05	-0.01	0.02	-0.06	-0.03	-0.01	0.03
	nPSO1 C=10	-0.20	-0.14	-0.06	-0.03	-0.04	-0.05	-0.02	0.02	-0.03	-0.02	0.00	0.03
	nPSO1 C=15	-0.19	-0.14	-0.06	-0.03	-0.05	-0.06	-0.02	0.02	-0.02	-0.02	0.00	0.03
	nPSO1 C=20	-0.19	-0.13	-0.07	-0.04	-0.06	-0.07	-0.02	0.02	-0.02	-0.03	-0.01	0.03
	nPSO2 C=5	-0.15	-0.09	-0.05	-0.03	-0.09	-0.04	-0.01	0.02	-0.07	-0.04	-0.01	0.02
	nPSO2 C=10	-0.18	-0.12	-0.05	-0.03	-0.07	-0.04	-0.01	0.02	-0.05	-0.02	0.00	0.04
	nPSO2 C=15	-0.18	-0.12	-0.05	-0.02	-0.05	-0.05	0.00	0.03	-0.04	-0.02	0.00	0.03
	nPSO2 C=20	-0.18	-0.12	-0.07	-0.03	-0.06	-0.05	-0.01	0.03	-0.04	-0.02	0.01	0.04
	nPSO3 C=5	-0.15	-0.11	-0.05	-0.03	-0.07	-0.05	-0.01	0.02	-0.06	-0.04	-0.01	0.02
	nPSO3 C=10	-0.20	-0.14	-0.07	-0.03	-0.05	-0.05	-0.02	0.02	-0.03	-0.03	-0.01	0.03
	nPSO3 C=15	-0.20	-0.13	-0.06	-0.03	-0.05	-0.06	-0.02	0.02	-0.03	-0.03	-0.01	0.03
	nPSO3 C=20	-0.19	-0.13	-0.07	-0.03	-0.06	-0.07	-0.02	0.02	-0.02	-0.03	-0.01	0.03
m=20	PSO	-0.20	-0.12	-0.07	-0.04	-0.10	-0.09	-0.05	0.00	-0.07	-0.07	-0.04	0.01
	nPSO1 C=5	-0.19	-0.12	-0.05	-0.03	-0.07	-0.05	-0.01	0.02	-0.07	-0.04	-0.01	0.02
	nPSO1 C=10	-0.22	-0.12	-0.06	-0.03	-0.06	-0.06	-0.02	0.02	-0.04	-0.03	-0.01	0.03
	nPSO1 C=15	-0.20	-0.13	-0.06	-0.03	-0.07	-0.08	-0.03	0.01	-0.03	-0.04	-0.01	0.03
	nPSO1 C=20	-0.21	-0.13	-0.07	-0.04	-0.08	-0.09	-0.03	0.02	-0.03	-0.05	-0.01	0.03
	nPSO2 C=5	-0.17	-0.09	-0.04	-0.03	-0.11	-0.05	-0.01	0.01	-0.08	-0.04	-0.01	0.02
	nPSO2 C=10	-0.19	-0.11	-0.05	-0.02	-0.08	-0.05	-0.01	0.02	-0.06	-0.03	0.00	0.03
	nPSO2 C=15	-0.20	-0.11	-0.06	-0.03	-0.07	-0.06	-0.01	0.02	-0.05	-0.03	0.00	0.03
	nPSO2 C=20	-0.20	-0.11	-0.06	-0.04	-0.08	-0.07	-0.02	0.02	-0.04	-0.03	0.00	0.03
	nPSO3 C=5	-0.19	-0.11	-0.06	-0.04	-0.07	-0.06	-0.02	0.02	-0.07	-0.04	-0.01	0.02
	nPSO3 C=10	-0.21	-0.13	-0.05	-0.03	-0.06	-0.07	-0.02	0.02	-0.04	-0.04	-0.01	0.02
	nPSO3 C=15	-0.20	-0.12	-0.06	-0.03	-0.08	-0.08	-0.03	0.01	-0.04	-0.04	-0.01	0.02
	nPSO3 C=20	-0.20	-0.12	-0.07	-0.04	-0.09	-0.09	-0.03	0.02	-0.04	-0.05	-0.02	0.03

Figure 8. Assortativity of the PSO and nPSO networks. Synthetic networks have been generated using the nPSO model with parameters $N = [100, 500, 1000]$, $m = [5, 10, 15, 20]$, $T = [0, 0.3, 0.6, 0.9]$, $\gamma = 3$ and angular coordinates sampled according to mixture distributions of three different kinds with components $C = [5, 10, 15, 20]$: Gaussian mixture with equal proportions (nPSO1), Gaussian mixture with random proportions (nPSO2), Gaussian and Gamma mixture with equal proportions (nPSO3). For more details on the parameters of the mixture distributions please refer to the Methods. Furthermore, synthetic networks have been generated also using the PSO model with the same parameters N, m, T and γ. For each combination of parameters, 10 networks have been generated and the assortativity has been computed. The heatmap reports for each parameter combination the mean assortativity, coloured according to a blue-to-red colormap in the range $[-0.3, 0.3]$.



		N=100				N=500				N=1000			
		T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9
m=5	PSO	0.99	0.94	0.92	0.91	1.00	0.92	0.87	0.87	1.00	0.92	0.85	0.84
	nPSO1 C=5	1.00	0.93	0.91	0.90	1.00	0.90	0.85	0.85	1.00	0.89	0.84	0.82
	nPSO1 C=10	1.00	0.94	0.91	0.92	1.00	0.91	0.86	0.85	1.00	0.90	0.83	0.83
	nPSO1 C=15	1.00	0.94	0.91	0.90	1.00	0.91	0.85	0.85	1.00	0.90	0.83	0.83
	nPSO1 C=20	1.00	0.95	0.91	0.91	1.00	0.91	0.86	0.85	1.00	0.91	0.84	0.82
	nPSO2 C=5	1.00	0.93	0.90	0.90	1.00	0.89	0.86	0.86	1.00	0.88	0.83	0.84
	nPSO2 C=10	1.00	0.93	0.91	0.91	1.00	0.89	0.86	0.85	1.00	0.88	0.83	0.83
	nPSO2 C=15	1.00	0.94	0.90	0.90	1.00	0.90	0.86	0.85	1.00	0.89	0.82	0.82
	nPSO2 C=20	1.00	0.94	0.91	0.91	1.00	0.90	0.85	0.85	1.00	0.89	0.83	0.82
	nPSO3 C=5	1.00	0.93	0.91	0.91	1.00	0.91	0.85	0.86	1.00	0.89	0.83	0.83
	nPSO3 C=10	1.00	0.94	0.91	0.91	1.00	0.91	0.85	0.86	1.00	0.90	0.83	0.83
	nPSO3 C=15	1.00	0.94	0.91	0.91	1.00	0.92	0.87	0.85	1.00	0.90	0.84	0.83
	nPSO3 C=20	1.00	0.94	0.92	0.92	1.00	0.91	0.87	0.85	1.00	0.91	0.85	0.83
m=10	PSO	1.00	0.97	0.95	0.95	1.00	0.96	0.92	0.90	1.00	0.96	0.90	0.88
	nPSO1 C=5	1.00	0.97	0.95	0.95	1.00	0.95	0.91	0.91	1.00	0.94	0.89	0.89
	nPSO1 C=10	1.00	0.97	0.95	0.95	1.00	0.95	0.91	0.91	1.00	0.95	0.90	0.89
	nPSO1 C=15	1.00	0.97	0.95	0.95	1.00	0.96	0.92	0.91	1.00	0.95	0.90	0.89
	nPSO1 C=20	1.00	0.97	0.95	0.95	1.00	0.96	0.92	0.91	1.00	0.95	0.90	0.89
	nPSO2 C=5	1.00	0.96	0.95	0.95	1.00	0.93	0.91	0.91	1.00	0.92	0.89	0.88
	nPSO2 C=10	1.00	0.97	0.95	0.95	1.00	0.94	0.91	0.91	1.00	0.93	0.89	0.89
	nPSO2 C=15	1.00	0.97	0.95	0.95	1.00	0.95	0.91	0.91	1.00	0.93	0.89	0.88
	nPSO2 C=20	1.00	0.97	0.95	0.95	1.00	0.95	0.91	0.91	1.00	0.94	0.90	0.89
	nPSO3 C=5	1.00	0.97	0.95	0.95	1.00	0.95	0.91	0.90	1.00	0.94	0.89	0.88
	nPSO3 C=10	1.00	0.97	0.96	0.95	1.00	0.95	0.92	0.91	1.00	0.95	0.90	0.88
	nPSO3 C=15	1.00	0.97	0.95	0.95	1.00	0.96	0.92	0.91	1.00	0.95	0.90	0.88
	nPSO3 C=20	1.00	0.97	0.95	0.95	1.00	0.96	0.92	0.91	1.00	0.95	0.90	0.89
m=15	PSO	1.00	0.98	0.97	0.96	1.00	0.97	0.93	0.92	1.00	0.97	0.92	0.90
	nPSO1 C=5	1.00	0.98	0.97	0.96	1.00	0.96	0.93	0.92	1.00	0.95	0.91	0.90
	nPSO1 C=10	1.00	0.98	0.97	0.96	1.00	0.97	0.93	0.92	1.00	0.96	0.91	0.90
	nPSO1 C=15	1.00	0.98	0.97	0.96	1.00	0.97	0.93	0.92	1.00	0.96	0.92	0.90
	nPSO1 C=20	1.00	0.98	0.97	0.96	1.00	0.97	0.93	0.92	1.00	0.96	0.92	0.90
	nPSO2 C=5	1.00	0.98	0.97	0.97	1.00	0.95	0.93	0.92	1.00	0.94	0.91	0.90
	nPSO2 C=10	1.00	0.98	0.97	0.96	1.00	0.96	0.93	0.92	1.00	0.94	0.91	0.90
	nPSO2 C=15	1.00	0.98	0.97	0.96	1.00	0.96	0.93	0.92	1.00	0.95	0.91	0.90
	nPSO2 C=20	1.00	0.98	0.97	0.97	1.00	0.96	0.93	0.92	1.00	0.95	0.91	0.90
	nPSO3 C=5	1.00	0.98	0.97	0.97	1.00	0.96	0.93	0.92	1.00	0.95	0.91	0.90
	nPSO3 C=10	1.00	0.98	0.97	0.96	1.00	0.97	0.93	0.92	1.00	0.96	0.91	0.90
	nPSO3 C=15	1.00	0.98	0.97	0.96	1.00	0.97	0.93	0.92	1.00	0.96	0.92	0.90
	nPSO3 C=20	1.00	0.98	0.97	0.96	1.00	0.97	0.93	0.92	1.00	0.97	0.92	0.90
m=20	PSO	1.00	0.98	0.98	0.97	1.00	0.97	0.94	0.93	1.00	0.97	0.93	0.91
	nPSO1 C=5	1.00	0.99	0.98	0.97	1.00	0.97	0.94	0.93	1.00	0.96	0.92	0.91
	nPSO1 C=10	1.00	0.98	0.98	0.97	1.00	0.97	0.94	0.93	1.00	0.96	0.93	0.91
	nPSO1 C=15	1.00	0.98	0.98	0.97	1.00	0.97	0.94	0.93	1.00	0.97	0.93	0.91
	nPSO1 C=20	1.00	0.98	0.98	0.97	1.00	0.97	0.94	0.93	1.00	0.97	0.93	0.91
	nPSO2 C=5	1.00	0.98	0.97	0.97	1.00	0.96	0.94	0.93	1.00	0.95	0.92	0.91
	nPSO2 C=10	1.00	0.98	0.98	0.97	1.00	0.96	0.94	0.93	1.00	0.95	0.92	0.91
	nPSO2 C=15	1.00	0.98	0.98	0.97	1.00	0.97	0.94	0.93	1.00	0.96	0.92	0.92
	nPSO2 C=20	1.00	0.98	0.98	0.97	1.00	0.97	0.94	0.93	1.00	0.96	0.93	0.92
	nPSO3 C=5	1.00	0.98	0.98	0.97	1.00	0.97	0.94	0.93	1.00	0.96	0.92	0.91
	nPSO3 C=10	1.00	0.98	0.98	0.97	1.00	0.97	0.94	0.93	1.00	0.97	0.93	0.91
	nPSO3 C=15	1.00	0.98	0.98	0.97	1.00	0.97	0.94	0.93	1.00	0.97	0.93	0.91
	nPSO3 C=20	1.00	0.98	0.98	0.97	1.00	0.97	0.94	0.93	1.00	0.97	0.93	0.91

Figure 9. LCP-correlation of the PSO and nPSO networks. Synthetic networks have been generated using the nPSO model with parameters $N = [100, 500, 1000]$, $m = [5, 10, 15, 20]$, $T = [0, 0.3, 0.6, 0.9]$, $\gamma = 3$ and angular coordinates sampled according to mixture distributions of three different kinds with components $C = [5, 10, 15, 20]$: Gaussian mixture with equal proportions (nPSO1), Gaussian mixture with random proportions (nPSO2), Gaussian and Gamma mixture with equal proportions (nPSO3). For more details on the parameters of the mixture distributions please refer to the Methods. Furthermore, synthetic networks have been generated also using the PSO model with the same parameters N, m, T and γ. For each combination of parameters, 10 networks have been generated and the LCP-correlation has been computed. The heatmap reports for each parameter combination the mean LCP-correlation, coloured according to a blue-to-red colormap in the range [0.5, 1].



		N=100				N=500				N=1000			
		T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9	T=0	T=0.3	T=0.6	T=0.9
m=5	PSO	0.50	0.29	0.22	0.20	0.43	0.20	0.11	0.08	0.41	0.17	0.09	0.06
	nPSO1 C=5	0.51	0.29	0.23	0.22	0.42	0.16	0.10	0.09	0.41	0.14	0.08	0.06
	nPSO1 C=10	0.51	0.30	0.23	0.22	0.43	0.17	0.11	0.09	0.41	0.14	0.08	0.07
	nPSO1 C=15	0.52	0.30	0.21	0.21	0.44	0.18	0.11	0.10	0.41	0.15	0.09	0.07
	nPSO1 C=20	0.52	0.30	0.21	0.20	0.44	0.18	0.12	0.09	0.41	0.15	0.09	0.07
	nPSO2 C=5	0.49	0.27	0.22	0.21	0.42	0.16	0.09	0.08	0.41	0.15	0.07	0.06
	nPSO2 C=10	0.51	0.29	0.23	0.22	0.43	0.16	0.11	0.09	0.41	0.14	0.08	0.06
	nPSO2 C=15	0.50	0.29	0.22	0.21	0.43	0.16	0.11	0.09	0.41	0.14	0.08	0.07
	nPSO2 C=20	0.52	0.29	0.23	0.22	0.43	0.17	0.11	0.09	0.42	0.14	0.09	0.07
	nPSO3 C=5	0.51	0.28	0.23	0.22	0.42	0.17	0.10	0.09	0.41	0.15	0.08	0.06
	nPSO3 C=10	0.50	0.31	0.23	0.20	0.43	0.18	0.11	0.09	0.41	0.15	0.09	0.07
	nPSO3 C=15	0.51	0.29	0.22	0.20	0.43	0.18	0.12	0.09	0.41	0.15	0.09	0.07
	nPSO3 C=20	0.51	0.29	0.22	0.20	0.43	0.18	0.12	0.09	0.42	0.15	0.09	0.07
m=10	PSO	0.62	0.31	0.26	0.26	0.54	0.21	0.13	0.12	0.52	0.19	0.12	0.09
	nPSO1 C=5	0.63	0.32	0.27	0.29	0.53	0.17	0.13	0.13	0.52	0.15	0.10	0.10
	nPSO1 C=10	0.64	0.31	0.26	0.27	0.55	0.18	0.14	0.14	0.53	0.16	0.11	0.10
	nPSO1 C=15	0.64	0.30	0.27	0.25	0.55	0.20	0.15	0.13	0.53	0.17	0.12	0.11
	nPSO1 C=20	0.63	0.31	0.26	0.26	0.55	0.21	0.15	0.13	0.54	0.17	0.12	0.11
	nPSO2 C=5	0.64	0.29	0.28	0.28	0.54	0.16	0.12	0.12	0.51	0.15	0.10	0.09
	nPSO2 C=10	0.63	0.30	0.28	0.28	0.54	0.18	0.14	0.13	0.52	0.16	0.11	0.10
	nPSO2 C=15	0.64	0.31	0.27	0.27	0.55	0.19	0.14	0.14	0.52	0.16	0.11	0.10
	nPSO2 C=20	0.64	0.31	0.27	0.27	0.55	0.19	0.14	0.14	0.53	0.16	0.11	0.11
	nPSO3 C=5	0.63	0.32	0.28	0.28	0.54	0.17	0.13	0.13	0.52	0.15	0.11	0.10
	nPSO3 C=10	0.64	0.32	0.27	0.26	0.55	0.19	0.14	0.13	0.52	0.16	0.11	0.10
	nPSO3 C=15	0.63	0.31	0.27	0.27	0.55	0.20	0.14	0.13	0.53	0.17	0.12	0.10
	nPSO3 C=20	0.62	0.31	0.27	0.26	0.55	0.21	0.15	0.13	0.54	0.17	0.12	0.11
m=15	PSO	0.72	0.35	0.32	0.31	0.60	0.22	0.16	0.14	0.59	0.20	0.13	0.11
	nPSO1 C=5	0.70	0.35	0.32	0.33	0.62	0.19	0.16	0.16	0.58	0.16	0.12	0.12
	nPSO1 C=10	0.74	0.35	0.32	0.32	0.62	0.21	0.17	0.17	0.60	0.17	0.13	0.12
	nPSO1 C=15	0.73	0.35	0.31	0.31	0.61	0.22	0.17	0.16	0.61	0.18	0.14	0.13
	nPSO1 C=20	0.71	0.34	0.32	0.31	0.60	0.23	0.17	0.16	0.61	0.19	0.14	0.13
	nPSO2 C=5	0.72	0.34	0.32	0.34	0.61	0.18	0.15	0.15	0.58	0.16	0.12	0.11
	nPSO2 C=10	0.73	0.34	0.32	0.33	0.61	0.20	0.16	0.16	0.59	0.17	0.12	0.12
	nPSO2 C=15	0.72	0.34	0.32	0.33	0.62	0.21	0.16	0.17	0.60	0.18	0.13	0.13
	nPSO2 C=20	0.72	0.35	0.32	0.33	0.61	0.21	0.16	0.17	0.60	0.18	0.13	0.13
	nPSO3 C=5	0.71	0.35	0.32	0.34	0.62	0.19	0.16	0.16	0.58	0.17	0.12	0.12
	nPSO3 C=10	0.74	0.35	0.32	0.31	0.61	0.21	0.16	0.16	0.60	0.18	0.13	0.12
	nPSO3 C=15	0.72	0.35	0.32	0.31	0.61	0.22	0.17	0.16	0.60	0.18	0.13	0.13
	nPSO3 C=20	0.72	0.35	0.31	0.32	0.60	0.23	0.17	0.16	0.61	0.19	0.14	0.13
m=20	PSO	0.78	0.38	0.36	0.36	0.64	0.24	0.18	0.17	0.63	0.21	0.14	0.13
	nPSO1 C=5	0.79	0.39	0.37	0.38	0.67	0.21	0.18	0.18	0.64	0.17	0.14	0.14
	nPSO1 C=10	0.79	0.38	0.36	0.36	0.66	0.23	0.19	0.19	0.65	0.19	0.15	0.15
	nPSO1 C=15	0.78	0.39	0.36	0.37	0.64	0.24	0.19	0.19	0.65	0.20	0.15	0.15
	nPSO1 C=20	0.77	0.38	0.36	0.36	0.64	0.24	0.18	0.18	0.64	0.21	0.15	0.15
	nPSO2 C=5	0.78	0.38	0.36	0.38	0.66	0.21	0.17	0.18	0.63	0.17	0.13	0.13
	nPSO2 C=10	0.79	0.39	0.36	0.38	0.66	0.22	0.18	0.19	0.64	0.18	0.14	0.14
	nPSO2 C=15	0.78	0.38	0.36	0.37	0.66	0.23	0.18	0.19	0.64	0.20	0.14	0.14
	nPSO2 C=20	0.78	0.38	0.37	0.37	0.66	0.23	0.18	0.19	0.64	0.19	0.15	0.15
	nPSO3 C=5	0.78	0.39	0.36	0.38	0.67	0.21	0.18	0.18	0.63	0.18	0.14	0.14
	nPSO3 C=10	0.78	0.38	0.36	0.36	0.65	0.23	0.18	0.19	0.65	0.19	0.15	0.14
	nPSO3 C=15	0.78	0.38	0.36	0.36	0.64	0.24	0.18	0.19	0.65	0.20	0.15	0.15
	nPSO3 C=20	0.77	0.38	0.36	0.36	0.65	0.24	0.18	0.18	0.64	0.21	0.15	0.15

Figure 10. Structural consistency of the PSO and nPSO networks. Synthetic networks have been generated using the nPSO model with parameters $N = [100, 500, 1000]$, $m = [5, 10, 15, 20]$, $T = [0, 0.3, 0.6, 0.9]$, $\gamma = 3$ and angular coordinates sampled according to mixture distributions of three different kinds with components $C = [5, 10, 15, 20]$: Gaussian mixture with equal proportions (nPSO1), Gaussian mixture with random proportions (nPSO2), Gaussian and Gamma mixture with equal proportions (nPSO3). For more details on the parameters of the mixture distributions please refer to the Methods. Furthermore, synthetic networks have been generated also using the PSO model with the same parameters N, m, T and γ. For each combination of parameters, 10 networks have been generated and the structural consistency has been computed. The heatmap reports for each parameter combination the mean structural consistency, coloured according to a blue-to-red colormap in the range [0, 1].

angular coordinate distribution (nPSO1, nPSO2 or nPSO3) and the number of communities (C) do not have a remarkable effect for most of the parameter combinations. The few cases in which there is higher variability are for $N = [500, 1000]$ and $T = 0.3$ where, with the increase of the number of communities, the clustering coefficient increases, becoming closer and closer to the one of the PSO model.

Figure 7 shows the characteristic path length, which describes the average of the shortest path lengths between all the pairs of vertices [13]. The measure decreases for increasing temperature, since there is higher probability to establish connections between nodes that are far apart from each other, acting as bridges between different, and often far apart, regions of the network. This decrease of the characteristic path length is attenuated when there are many edges with respect to the network size (higher m and lower N), because the bridges naturally emerge due to the high network density. Increasing m, indeed, leads in general to a decrease of the characteristic path length. The type of angular coordinate distribution (nPSO1, nPSO2 or nPSO3) and the number of communities (C) do not have any remarkable effect.

Figure 8 reports the assortativity, which indicates the tendency of the networks to connect nodes with similar degree [28]. Positive values suggest an assortative behaviour and negative values a disassortative mixing. The results highlight that there are no parameter combinations for which the networks are strongly assortative, whereas disassortativity is detected for networks of small size at low temperature. In the generative procedure of the PSO and nPSO models, the oldest nodes have the highest node degree and, being at the centre of the hyperbolic disk, at low temperature they tend to be the connection targets of new coming nodes with lower degree, leading to a disassortative mixing. At higher temperature this disassortativity gets weaker and there is more balance between the connections established from a new node to popular and less popular nodes, resulting in an increase of the measure. It can be noticed also a decrease at $N = 100$ and $T = 0$ for increasing m, since more connections are created between the older higher-degree nodes and the younger lower degree nodes. The type of angular coordinate distribution (nPSO1, nPSO2 or nPSO3) and the number of communities (C) lead only to minor variabilities for low temperature with respect to the PSO.

Figure 9 shows the LCP-correlation, the Pearson correlation between the number of common neighbours (which create a local community) and the number of local community links (connections among common neighbours) that are computed for each link of the network [29]. The LCP-correlation measures whether the network follows a LCP organization [29–31] and therefore whether the network is organized in local communities (one for each link) where the number of interactions between the common neighbours is a function that increases with the number of common neighbours in the local community. Complex adaptive networks with weak-links that make local processing and global delivery generally follow the LCP organization (the LCP-correlation is generally $\geqslant 0.7$), whereas the networks that do not follow the LCP organization (LCP-correlation $\leqslant 0.3$) present strong-links, they are not clustered and they are suitable for storage or mere delivery of energy or information. It is very rare to find networks that have a LCP-correlation between 0.3 and 0.7. For this reason we expect to find that all the nPSO networks are organized according to the LCP, however the level of LCP-correlation might change between 0.7 and 1. Indeed, as expected, the LCP-correlation obtains high values over all the parameter combinations of the network. Since connection probabilities depend on geometrical distances, for a given link of the network it is likely that the adjacent nodes are close in the hyperbolic space, therefore it is likely that their common neighbours (if any) are close, and as a consequence it is also likely that these common neighbours have connections among them that increase with the number of common neighbours. Explained in a simpler form: the smaller the hyperbolic distance between two linked nodes, the more common neighbours exist between them (since the geometrical space that separates the two linked points is smaller, the two linked points share more adjacent nodes, which are in fact common neighbours), as a consequence the smaller geometrical space will generate also more connections between these common neighbours. This mechanism obviously is corrupted for increasing temperature, since the connection probabilities have a weaker dependency on the geometrical distances, and therefore the LCP-correlation decreases for high temperatures. In particular, this temperature-dependent LCP-correlation decrease is remarkable for lower m and higher N, since there are less links to establish and more possible connection targets, which reduces the probability to create both common neighbours and local community links (links between the common neighbours). The type of angular coordinate distribution (nPSO1, nPSO2 or nPSO3) and the number of communities (C) lead only to minor variabilities for $T > 0$.

Figure 10 reports the structural consistency, which quantifies the link predictability of the network, characterizing the inherent difficulty to predict the missing or non-observed links regardless of the specific algorithm used for the prediction [32]. The structural consistency strongly decreases for increasing temperature, in particular from $T = 0$ to $T > 0$. In fact, at $T = 0$ the links are regularly established with the closest target nodes, which makes the structure highly consistent and easier to predict. Furthermore, creating a higher number of connections according to this regular pattern ($T = 0$ and higher m) strengthens even more the consistency of the structure. The link predictability becomes lower for increasing network size, since there are potentially more missing or non-observed links to predict. The type of angular coordinate distribution (nPSO1, nPSO2 or nPSO3) and the number of communities (C) lead only to minor variabilities.

Supplementary figure 9 shows the exponent γ of the power-law degree distribution, fitted using the procedure described by Clauset *et al* [33], in order to test whether the value provided in input to the PSO and nPSO models is indeed reproduced. The results highlight that all the fitted values are very close to the desired exponent $\gamma = 3$. The variability might be either due to the difficulty of the model to reproduce perfectly the input value or due to some defects in the fitting procedure. The diverse community organization does not introduce a remarkable bias in the degree distribution.

Supplementary figure 10 reports the modularity, indicating the extent to which the network can be partitioned in segregated modules that tend to interact densely within themselves but sparsely between each other [17]. We let notice that in nPSO networks the modularity is inversely related to the community mixing, since the lower the community mixing the more the network can be partitioned in distinct modules. Indeed, the Pearson correlation between the community mixing and the modularity over all the parameter combinations of the nPSO model is -0.91. The main patterns observed on nPSO networks for the community mixing are therefore valid, in an inverse way, also for the modularity. For PSO networks the modularity is generally lower, with an exception for larger networks, low m and low temperature, probably due to the fact that many small modules naturally emerge since for low temperatures the clustering is very high. The type of angular coordinate distribution (nPSO1, nPSO2 or nPSO3) lead only to minor variabilities, although we let notice that such variabilities might be even due to the randomness in the modularity evaluation procedure.

Supplementary figure 11 shows the measure of small-worldness ω, which indicates whether a network exhibits a small-world organization, characterized by a clustering coefficient (CL) as high as in an equivalent lattice network ($\mathrm{CL}_{\mathrm{latt}}$) and a characteristic path length (L) as low as in an equivalent random network (L_{rand}) [13, 34]: $\omega = \frac{L_{\mathrm{rand}}}{L} - \frac{\mathrm{CL}}{\mathrm{CL}_{\mathrm{latt}}}$. The measure ω is expected to be close to 0 in small-world networks ($L \approx L_{\mathrm{rand}}$ and $\mathrm{CL} \approx \mathrm{CL}_{\mathrm{latt}}$), higher than 0 for random networks ($L \approx L_{\mathrm{rand}}$ and $\mathrm{CL} < \mathrm{CL}_{\mathrm{latt}}$) and lower than 0 for lattice networks ($L > L_{\mathrm{rand}}$ and $\mathrm{CL} \approx \mathrm{CL}_{\mathrm{latt}}$). The parameter combinations closest to small-world networks are at $N = 100$, $m = [15, 20]$ and $T = 0$. Indeed, these are the synthetic networks characterized by the highest clustering coefficient and the lowest characteristic path length. The measure ω increases for increasing temperature, since the clustering coefficient strongly decreases and the characteristic path length slightly decreases, with a transition from structural properties of a regular network to the ones of a random network. At $T = 0$, the measure ω increases for increasing m, since the clustering coefficient is constantly high and the characteristic path length decreases. This is not always valid at $T > 0$, where sometimes the increase in clustering balances the decrease of the characteristic path length. The type of angular coordinate distribution (nPSO1, nPSO2 or nPSO3) and the number of communities (C) lead only to minor variabilities.

Supplementary figure 12 reports the p-value of the statistical test for rich-clubness, which indicates whether the network presents a significant rich-club organization with respect to the Cannistraci–Muscoloni null model [35]. It clearly emerges that for almost all the parameter combinations the synthetic networks are significantly (p-value $\leqslant 0.05$) characterized by rich-clubness. This is indeed in agreement with the network growing procedure explained by the PSO and nPSO models. In fact, the high degree nodes are the first ones to be born in the network and they connect to m of the nodes already present [5], therefore every network has at least a fully connected subgraph composed by the $m + 1$ oldest high degree nodes. The only p-values that are not significant are borderline and detected for small ($N = 100$) and sparse ($m = 5$) networks, since there are less nodes and links to build the rich-club, and only at higher temperature, where the connection probabilities have a weaker dependency from geometrical distances and therefore the rich and popular nodes decrease their attractiveness for new connections. The type of angular coordinate distribution (nPSO1, nPSO2 or nPSO3) and the number of communities (C) lead only to minor variabilities among the borderline cases.

All these topological measures have been evaluated using the MATLAB code released at: https://github.com/biomedical-cybernetics/topological_measures_wide_analysis [36].

Conclusion

Recent studies presented the hyperbolic disk as an adequate space to describe the latent geometry of real complex networks and the PSO model was introduced to generate random geometric graphs in the hyperbolic space, reproducing strong clustering and a scale-free degree distribution [5]. Coupling the hyperbolic space with the preferential attachment of nodes to this space, the GPA model confers to the networks also a community structure, introducing the idea that different angular regions of the hyperbolic disk can have a variable level of attractiveness [19]. However, the GPA model does not allow to indicate in input a desired number of communities, neither to control their size and the mixing between them, which is a clear limitation for real applications. For this reason, we here introduced the nPSO model, which allows one to explicitly fix the number of communities and their size by means of a tailored probability distribution on the angular coordinates, and to tune the mixing property through the network temperature.

We performed extensive tests on the detectability of the nPSO communities, considering also more complicated settings with asymmetric angular coordinate distributions over the angular space. We highlighted that, for most of the parameter combinations representing realistic scenarios, the community organization can be spotted by the state-of-the-art algorithm Louvain. The main factor that reduces the detectability is the ratio between the number of communities and the network size, in particular community detection in nPSO networks reduces significantly for small size networks that present many communities. These results suggest that realistic community structure is properly reproduced by the model and the nPSO might be employed in future studies as a benchmark for testing community detection algorithms. On this regard, we propose a second study that discusses how to leverage the nPSO model to test and compare the performance of different algorithms for community detection and also link prediction [26].

We evaluated and compared several topological measures of the synthetic networks generated using the PSO and nPSO models, and from this wide investigation two important results emerge. First, the parameters of the model allow to reproduce a great variety of the structural properties observed in real-world complex networks, and the heatmaps provided in this study can be used as a reference for the choice of parameters while generating networks with desired characteristics. Second, the diverse community organization has only a minor impact for most of the main topological measures. This suggests, for example, that the temperature of a real network can be inferred from the clustering coefficient regardless of the community structure.

From the algorithmic point of view, since the original procedure to establish links adopted by the PSO and GPA models is computationally expensive for generating networks with communities and high clustering, we proposed other two different variants. We demonstrated that the three implementations generate equivalent topologies and the fastest of them (implementation 3) significantly reduces the computational time, with a complexity of $O(EN)$ independently from the communities and the clustering.

Although in this work we present the nPSO as a generative model for realistic networks with non-overlapping communities, its current implementation would be able also to generate networks with overlapping communities, for instance by increasing the standard deviations of the components in the Gaussian mixture distribution. However, a specific rule according to which the nodes are assigned to one or more communities needs to be designed, depending both on the geometrical positions of the nodes (angular and radial coordinates) and on the mixture distribution parameters. This extension of the nPSO model will be investigated in future studies.

To conclude, we propose the nPSO model as a valid framework able to efficiently generate realistic networks with a fixed number of communities according to a nonuniform node-angular probability distribution. The nPSO might be adopted, among the many possibilities, as a null model for the hyperbolic embedding of networks with community structure, or as a benchmark for testing community detection and link prediction algorithms, as we illustrate and discuss in a second study dedicated to this topic [26].

Code availability

The MATLAB code for generating synthetic networks using the nPSO model is publicly available at the GitHub repository:

https://github.com/biomedical-cybernetics/nPSO_model

Hardware and software

MATLAB code has been used for all the simulations, carried out partly on a workstation under Windows 8.1 Pro with 512 GB of RAM and 2 Intel(R) Xenon(R) CPU E5-2687W v3 processors with 3.10 GHz, and partly in the ZIH-Cluster Taurus of the TU Dresden.

Funding

Work in the CVC laboratory was supported by the independent research group leader starting grant of the Technische Universität Dresden. AM was partially supported by the funding provided by the Free State of Saxony in accordance with the Saxon Scholarship Program Regulation, awarded by the Studentenwerk Dresden based on the recommendation of the board of the Graduate Academy of TU Dresden. We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the TU Dresden.

Acknowledgments

We thank Alexander Mestiashvili and the BIOTEC System Administrators for their IT support, Claudia Matthes for the administrative assistance and the Centre for Information Services and High Performance Computing (ZIH) of the TUD.

Author contributions

CVC invented the nPSO model and designed the numerical experiments. AM implemented the code and performed the computational analysis. Both the authors analysed and interpreted the results. AM and CVC built the demonstration of the equivalence of the three implementations for link generation and AM formalized it. AM performed the analysis for the computational complexity and CVC checked it. AM wrote the draft of the article according to CVC suggestions and CVC corrected and improved it to arrive to the final draft. CVC designed the figures and AM realized them. AM designed and realized the heatmap tables. CVC planned, directed and supervised the study.

Competing interests

The authors declare no competing financial interests.

ORCID iDs

Alessandro Muscoloni https://orcid.org/0000-0002-9238-3357
Carlo Vittorio Cannistraci https://orcid.org/0000-0003-0100-8410

References

[1] Serrano M Á, Krioukov D and Boguñá M 2008 Self-similarity of complex networks and hidden metric spaces *Phys. Rev. Lett.* **100** 078701

[2] Krioukov D, Papadopoulos F, Vahdat A and Boguñá M 2009 Curvature and temperature of complex networks *Phys. Rev.* E **80** 035101

[3] Krioukov D, Papadopoulos F, Kitsak M, Vahdat A and Boguñá M 2010 Hyperbolic geometry of complex networks *Phys. Rev.* E **82** 036106

[4] Boguñá M, Papadopoulos F and Krioukov D 2010 Sustaining the Internet with hyperbolic mapping *Nat. Commun.* **1** 1–8

[5] Papadopoulos F, Kitsak M, Serrano M Á, Boguñá M and Krioukov D 2012 Popularity versus similarity in growing networks *Nature* **489** 537–40

[6] Kleineberg K-K, Boguñá M, Serrano M Á and Papadopoulos F 2016 Hidden geometric correlations in real multiplex networks *Nat. Phys.* **12** 1076–81

[7] Bianconi G and Rahmede C 2017 Emergent hyperbolic network geometry *Sci. Rep.* **7** 41974

[8] Allard A, Serrano M Á, García-Pérez G and Boguñá M 2016 The geometric nature of weights in real complex networks *Nat. Commun.* **8** 14103

[9] Muscoloni A, Thomas J M, Ciucci S, Bianconi G and Cannistraci C V 2017 Machine learning meets complex networks via coalescent embedding in the hyperbolic space *Nat. Commun.* **8** 1615

[10] Muscoloni A and Cannistraci C V 2018 Minimum curvilinear automata with similarity attachment for network embedding and link prediction in the hyperbolic space arXiv:1802.01183

[11] Cacciola A *et al* 2017 Coalescent embedding in the hyperbolic space unsupervisedly discloses the hidden geometry of the brain arXiv:1705.04192

[12] Muscoloni A and Cannistraci C V 2017 Local-ring network automata and the impact of hyperbolic geometry in complex network link-prediction arXiv:1707.09496

[13] Watts D J and Strogatz S H 1998 Collective dynamics of 'small-world' networks *Nature* **393** 440–2

[14] Barabasi A L and Albert R 1999 Emergence of scaling in random networks *Science* **286** 509–12

[15] Barabasi A L 2009 Scale-free networks: a decade and beyond *Science* **325** 412–3

[16] Girvan M and Newman M E J 2002 Community structure in social and biological networks *Proc. Natl Acad. Sci.* **99** 7821–6

[17] Newman M E J 2006 Modularity and community structure in networks *Proc. Natl Acad. Sci. USA* **103** 8577–82

[18] Fortunato S and Hric D 2016 Community detection in networks: a user guide *Phys. Rep.* **659** 1–44

[19] Zuev K, Boguñá M, Bianconi G and Krioukov D 2015 Emergence of soft communities from geometric preferential attachment *Sci. Rep.* **5** 9421

[20] McLachlan G and Peel D 2000 *Finite Mixture Models* (New York: Wiley)

[21] Danon L, Díaz-Guilera A, Duch J and Arenas A 2005 Comparing community structure identification *J. Stat. Mech.* P09008

[22] Blondel V D, Guillaume J-L, Lambiotte R and Lefebvre E 2008 Fast unfolding of communities in large networks *J. Stat. Mech.* 10008

[23] Yang Z, Algesheimer R and Tessone C J 2016 A comparative analysis of community detection algorithms on artificial networks *Sci. Rep.* **6** 30750

[24] Lancichinetti A and Fortunato S 2009 Community detection algorithms: a comparative analysis *Phys. Rev.* E **80** 56117

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

[26] Muscoloni A and Cannistraci C V 2018 Leveraging the nonuniform PSO network model as a benchmark for performance evaluation in community detection and link prediction *New J. Phys.* **20**

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

[28] Newman M E J 2002 Assortative mixing in networks *Phys. Rev. Lett.* **89** 208701

[29] Cannistraci C V, Alanis-Lobato G and Ravasi T 2013 From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks *Sci. Rep.* **3** 1–13

[30] Daminelli S, Thomas J M, Durán C and Cannistraci C V 2015 Common neighbours and the local-community-paradigm for topological link prediction in bipartite networks *New J. Phys.* **17** 113037

[31] Durán C, Daminelli S, Thomas J M, Haupt V J, Schroeder M and Cannistraci C V 2017 Pioneering topological methods for network-based drug–target prediction by exploiting a brain-network self-organization theory *Brief. Bioinform.* **8** 3–62

[32] Lü L, Pan L, Zhou T, Zhang Y-C and Stanley H E 2015 Toward link predictability of complex networks *Proc. Natl Acad. Sci.* **112** 2325–30

[33] Clauset A, Rohilla Shalizi C and Newman M E J 2009 Power-law distributions in empirical data *SIAM Rev.* **51** 661–703

[34] Telesford Q K, Joyce K E, Hayasaka S, Burdette J H and Laurienti P J 2011 The ubiquity of small-world networks *Brain Connect.* **1** 367–75

[35] Muscoloni A and Cannistraci C V 2017 Rich-clubness test: how to determine whether a complex network has or does not have a rich-club? arXiv:1704.03526

[36] Narula V, Zippo A G, Muscoloni A, Biella G E M and Cannistraci C V 2017 Can local-community-paradigm and epitopological learning enhance our understanding of how local brain connectivity is able to process, learn and memorize chronic pain? *Appl. Netw. Sci.* **2** 28