---

**Article:**
Wang, Jianjia and Hancock, Edwin R orcid.org/0000-0003-4496-2028 (2023) Statistical Structural Inference from Edge Weights using a Mixture of Gamma Distributions. Journal of Complex Networks. cnad038. ISSN 2051-1329

https://doi.org/10.1093/comnet/cnad038

---

*IMA Journal of Complex Networks* (2023) Page 1 of 24
doi:10.1093/comnet/xxx000

# Statistical Structural Inference from Edge Weights using a Mixture of Gamma Distributions

JIANJIA WANG*
*School of AI and Advanced Computing, XJTLU Entrepreneur College, Xi'an Jiaotong-Liverpool University, Suzhou, China, 215412.*
*Corresponding author: Jianjia.Wang@xjtlu.edu.cn

EDWIN R. HANCOCK
*Department of Computer Science, University of York, UK.*

[Received on 19 September 2023]

The inference of reliable and meaningful connectivity information from weights representing the affinity between nodes in a graph is an outstanding problem in network science. Usually, this is achieved by simply thresholding the edge weights to distinguish true links from false ones and to obtain a sparse set of connections. Tools developed in statistical mechanics have provided particularly effective ways to locate the optimal threshold so as to preserve the statistical properties of the network structure. Thermodynamic analogies together with statistical mechanical ensembles have been proven to be useful in analysing edge-weighted networks. To extend this work, in this paper, we use a statistical mechanical model to describe the probability distribution for edge weights. This models the distribution of edge weights using a mixture of Gamma distributions. Using a two-component Gamma mixture model with components describing the edge and non-edge weight distributions, we use the Expectation-Maximization algorithm to estimate the corresponding Gamma distribution parameters and mixing proportions. This gives the optimal threshold to convert weighted networks to sets of binary-valued connections. Numerical analysis shows that it provides a new way to describe the edge weight probability. Furthermore, using a physical analogy in which the weights are the energies of molecules in a solid, the probability density function for nodes is identical to the degree distribution resulting from a uniform weight on edges. This provides an alternative way to study the degree distribution with the nodal probability function in unweighted networks. We observe a phase transition in the low-temperature region, corresponding to a structural transition caused by applying the threshold. Experimental results on real-world weighted and unweighted networks reveal an improved performance for inferring binary edge connections from edge weights.

*Keywords*: structural inference, edge weights distribution, statistical mechanical model.

2000 Math Subject Classification: 34K30, 35K57, 35Q80, 92D25

## 1. Introduction

Most systems of interacting objects can be represented as networks and their study has attracted intense interest in analysing topological patterns. This usually involves aggregating structural or functional connections into an unweighted or weighted adjacency matrix in complex networks [1]. Much of the literature focuses on the statistical nature of structural patterns for unweighted networks, such as the power-law degree distribution for preferential attachment [2, 3]. Rather than representing an unweighted network as a set of binary connections, most real-world data sets contain fine-grained information for the strength of connections between each pair of vertices [4–6]. This is usually represented as a set of

edge weights, but this rarely considers statistical properties, such as the detailed weight distribution, in network structure.

Recently, there have been an increasing number of studies that commence by using the representation of real-world data as weighted networks and performing inference to transform them into a set of binary connections [7]. This provides the probabilistic estimation of the binary states (0 or 1) in the node adjacency matrix. In this form, it is particularly useful in brain imaging to remove inconsistent or weak connectivity in neuro-anatomical brain regions or drug design using protein-to-protein interaction networks. In both examples this is achieved by thresholding the given weight matrices to give a set of binary elements [8, 9]. This raises the controversial question concerning the optimal way to estimate or infer the underlying binary structure from weighted networks without losing potentially useful information [10].

One simple method for converting the edge weights into conventional binary connections is by thresholding with a fixed global value [11]. This weight-based method identifies those edges that exceed a certain constant edge weight value. The networks generated in this way usually have a variable number of edges, resulting in variations in the density of binary adjacency matrix elements [12]. An alternative way of thresholding is to retain a constant fraction of the strongest connections. This is a density-based thresholding method with a fixed number of edge connections. However, thresholding in this way can mask subtle variations in network topology, with a resulting loss of information conveyed by the patterns of edge density in individual networks. [5].

While these two different thresholding strategies can certainly reveal the underlying network structure, they have some limitations that highlight an important dilemma. The variations in the number of edges or edge density can affect the network topology but this poses problems when transforming weighted networks into binary connections [13]. A possible alternative is to retain the structural connections using statistical inference [14, 15]. Both unweighted and weighted networks are highly correlated structures. Both of them contain component edges and nodes, and the distribution of nodal degree combined with the distribution of edge weights can be used to represent topological patterns in the networks [7].

Although it is well-known that scale-free networks present a power-law node degree distribution, for scale-free networks, their statistical characterisation remains less certain. A recent study reveals that the power-law only fits well in the tail with an exponential cutoff [2]. This suggests a way to convert weighted networks to binary edge indicators. Provided with a suitable delineation of the two distributions, edges can be separated from non-edges via a simple thresholding strategy, thus reducing the redundancy in a heavy-tailed distribution [16].

To improve the performance of the thresholding of weighted networks to binary connections, the work described in this paper aims to establish an effective statistical method for describing the probability density function of nodes in weighted networks. We introduce a new way to calculate the degree distribution in terms of edge weights using an analogy in which the network nodes are particles in a solid [17]. The resulting distribution of edge weights can be approximated as a Gamma probability density function. This can be further represented as a mixture of two separate Gamma distributions for the edge and non-edge states, and their distribution parameters together with their mixing proportions can be estimated using the Expectation-Maximization algorithm. Using this method, we can find the optimal value of the threshold for edge weights in a network. This gives a statistical structural inference procedure for converting a weighted network to a binary adjacency matrix.

## 2. Related Work

***Thresholding.*** There is a rich literature on applying a threshold to the weighted network to distinguish real connections from spurious ones. To remove inconsistent or weak interactions, we assume that edges exist when the value of the weight exceeds a certain threshold which generates matrices with binary elements. Usually, there are two practical approaches to finding the optimal threshold. One method depends on the value of edge weights. It seeks a constant value to construct a variable number of edges in the network [18, 19]. An alternative method for thresholding is to retain a constant fraction of the strongest connections. This produces a fixed number of edges in the networks [20]. Instead of using these two global thresholding, some methods introduce local thresholding to avoid the problem of network fragmentation caused by a fixed value of the threshold. This computes thresholds locally at the node level, rather than over the entire network [21]. However, local thresholding can also create non-trivial topological structural artefacts which remains a problem in the identification of reliable network connections. Additional tools, such as integration over a range of thresholds have been introduced to avoid the arbitrariness in the choice of threshold [22, 23].

***Statistical Inference.*** There are three main approaches to the statistical analysis of network measures, a) omnibus testing, b) mass univariate testing and c) multivariate approaches. Omnibus testing is the simplest way to infer the network structure. This involves inference on one or more of the global topological measures, such as the average path length and the mean clustering coefficients [24]. The method applies standard statistical tests to detect the network structure with the desired characteristics. For example, permutation tests use the observed network structure to estimate the null distribution empirically, which provides a route to statistical inference where the real distribution of clustering coefficients is unknown [25]. Cantwell et al. propose a model for correlated relational data and explore the properties of the network ensemble by thresholding edge weights [12]. This is a simple way to obtain insights into global network properties but lacks specificity and is confined to a specific subset of nodes or edges.

Mass univariate testing provides a localization based on specific nodes or edges and goes beyond the use of global metrics. It applies a particular statistical test, such as a t-test, independently across a large number of nodes or connections. Usually, it focuses on a subset of connections or nodes and then performs inference on a specific node-specific measure [26]. When used to determine connections, it often tests connectivity for variations, such as edge weights or betweenness centrality. For example, Martin et al. propose a principled maximum-likelihood method for inferring community structure. They estimate the structure of the network from uncertain edge connections [14]. Mass univariate testing yields a statistical test to reject the null hypothesis at the level of individual nodes, connections or subnetworks. It allows inference based on local connectome elements or subnetworks [27].

Finally, multivariate approaches are usually combined with machine learning methods for statistical inference. They seek to recognize and learn the statistical structural patterns among multiple connections and utilize these patterns for inferential classification or prediction [28]. This includes a broad range of algorithms in pattern recognition, machine learning and deep learning [29]. For example, graph neural networks are proposed to learn both global topological structure and the local connectivity structure within a network [30]. Other algorithms, such as support vector machines, principal component analysis, or entropic models are also commonly used to perform statistical inference of network structure [31, 32].

## 3. Statistical Description of Weighted Networks

Let a weighted network $G(V,E,\omega)$ with node set $V$ and edge set $E \subseteq V \times V$. The symbol $\omega$ is the edge weight which means the pair of nodes $(u,v)$ contains a real non-negative value $w(u,v)$ for each edge,

i.e., $u \in V, v \in V$, and $u \neq v$.

The adjacency matrix of a graph is $A$ with the degree of node $u$ is $d_u = \sum_{v \in V} A_{uv}$. Then, the Laplacian matrix is $L = D - A$, where $D$ denotes the degree diagonal matrix whose elements are given by $D(u,u) = d_u$ and zeros elsewhere.

$$A_w = \begin{cases} w(u,v) & \text{if } (u,v) \in E \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

where, for the undirected network, the weighted adjacency is symmetric, i.e., $w(u,v) = w(v,u)$ for all pairs of nodes that $(u,v) \in E, u \neq v$.

We first define the total energy in this network as the summation of all edge weights,

$$U = \sum_{i=1}^{|E|} \omega_i = w|E| \tag{3.2}$$

This takes on an integer value if we assume all edge weights are unity, where $w = 1$ .

To compute the corresponding network entropy, we need to determine the number combinations in choosing $|E|$ edges among the available $|E| + |V| - 1$ possibilities [1]. This gives the entropy in terms of a combinatorial expression involving factorials as,

$$W(U,V) = \frac{(U + |V| - 1)!}{U!(|V| - 1)!} \tag{3.3}$$

For large networks, the entropy can be further simplified using Stirling's approximation $\log n! \approx n \log n - n$ with the logarithm of $W(U,V)$. It gives

$$\begin{aligned} S &= k_B \ln W \\ &= k_B \log[(U + |V| - 1)!] - \log(U!) - \log[(|V| - 1)!] \\ &\approx k_B[(U + |V| - 1)\log(U + |V| - 1) - U \log U - (|V| - 1)\log(|V| - 1)] \end{aligned} \tag{3.4}$$

where $k_B$ is the Boltzmann constant. To simplify the calculation, we set the Boltzmann constant to unity.

Finally, we compute the partial derivative of the entropy to energy at fixed the number of nodes in the graph. This derives the parameter $\beta$, which is called the inverse temperature. This is related to the rate of change of energy for entropy in the network.

$$\beta = \left( \frac{\partial S}{\partial U} \right)_{|V|} = \frac{1}{w} \log \frac{U + |V| - 1}{U} \tag{3.5}$$

where $w$ is the edge weight variable.

The inverse temperature $\beta = 1/T$ defined above implies that the temperature $T$ is proportional to the edge density in the network.

Here, the description of the continuous distribution in edge weights is given by a function $g(\omega)$ which is the density of edge weights. The edge weight in the infinitesimal interval between $\omega$ and $\omega + d\omega$ is given by $g(\omega)d\omega$ and the total number of edge weights is

$$\int g(\omega) d\omega = U \tag{3.6}$$

Eq.(3.2) defines the total energy to be the summation of edge weights. Here, for each edge, the weight is distributed over the two dimensions defined by the constituent node degrees of the edge. This can be extended to a more general case of the directed network with the in-degree and out-degree of the nodes.

As the nodal degree variables can be represented by a point on a two-dimensional degree-space, and these points are discrete and uniformly distributed in in-degree and out-degree dimensions. Each dimension is quantised by increments of unit degree. Single nodal degree-vectors whose magnitude lies between $k$ and $k+dk$ lies in one quadrant of a cycle with a radius $k$ and thickness $dk$. This quadrant degree-space only allows positive degree values.

Commencing from the degree distribution, we can replace the sum over two component edge degree vectors in Eq.(3.2) with an integral over the volume element $dk$. In this way, the discrete summation can be rewritten as an integral as

$$\sum_{k}(\cdots) = \frac{1}{4}\int_{0}^{\infty} 2\pi k(\cdots)dk \tag{3.7}$$

The number of nodes with a degree vector whose magnitude lies between $k$ and $k+dk$ can be described by the function $g(k)dk$. This gives the degree density per node as the area in the degree-space of one quadrant of a cycle surface divided by the total area in degree-space which can be occupied nodal degree vectors, i.e.

$$g(k)dk = \frac{|V|^2}{(2\pi)^2}\cdot 2\pi k\cdot dk\times 2 = \frac{|V|^2 k}{\pi}dk \tag{3.8}$$

where $|V|^2$ is the number of potential edges in the network, and the factor 2 represents the two possible edge dimensions.

The edge weight variable is related to the nodal degree multiplied by a unit of degree increment $\varepsilon$, i.e. $\omega = \varepsilon k$. Substituting into Eq.(3.8), this gives the density of weights for each node as

$$g(\omega)d\omega = \frac{|V|^2}{\pi\varepsilon^2}\omega d\omega \tag{3.9}$$

From statistical mechanics, we use the partition function to derive thermal quantities [33] for the network. Here, the logarithm of the partition function is given according to the density of weights as follows,

$$\log Z = \int_{0}^{\omega_T} g(\omega)d\omega \log\left[\frac{1}{1-e^{-\beta w}}\right] = -\int_{0}^{\omega_T} g(\omega)d\omega \log\left[1-e^{-\beta w}\right] \tag{3.10}$$

where $\omega_T$ is the upper limit boundary for the edge weights. The details will discussed later in Section 6.

Then, the average energy in a network is the partial derivative of the partition function to the inverse temperature,

$$\bar{U} = -\frac{\partial \log Z}{\partial \beta} = \int_{0}^{\omega_T} g(\omega)d\omega\cdot\frac{\omega}{e^{\beta\omega}-1} = \frac{S}{\pi\varepsilon^2}\int_{0}^{\omega_T}\frac{\omega^2}{e^{\beta\omega}-1}d\omega \tag{3.11}$$

The probability distribution of edge weights is given as

$$p(\omega) = \frac{\omega^2}{e^{\beta\omega}-1} \tag{3.12}$$

This probability density function is given in closed form by a Gamma distribution. It can be further approximated as the combination of Gamma functions. To convert the weighted network to an unweighted one, we assume there are two separate Gamma distribution functions that decompose this expression into edge and non-edge components. This mixture of Gamma distributions provides a route to computing the optimal value of the edge weight threshold [34].

Let $\omega = \{\omega_i\}, 1 \leqslant i \leqslant N$ be a set of edge weights in a network. These weights are regarded as independent and identical distribution random variables [31]. Eq.(3.12) considers that binary combinations of Gamma distribution as the mixture

$$p(\omega_i \mid \alpha, \beta) = \sum_{j=0}^{j=1} \pi_j G_W(\omega_i \mid \alpha_j, \beta_j) \tag{3.13}$$

where $\pi_j$ is the probability of the separate Gamma distribution. The condition $\sum_{j=1}^{j=1} \pi_j = 1$ must hold to guarantee that $p(\omega_i \mid \alpha, \beta)$ is a well-defined probability distribution. $\alpha_j$ and $\beta_j$ are the parameters of the component Gamma distributions. Thus, $G_W(.)$ i..e the Gamma probability density function is given as

$$G_W(\omega \mid \alpha, \beta) = \frac{\omega^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} e^{-(\omega/\beta)}, \omega \geqslant 0, \alpha, \beta > 0 \tag{3.14}$$

where $\Gamma(\alpha)$ is the Euler Gamma function that is $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$, for $\alpha \geqslant 0$.

Then, the joint distribution of edge weights is

$$p(\mathbf{W} \mid \Theta) = \prod_{i=1}^{|E|} p(\omega_i \mid \alpha, \beta) \tag{3.15}$$

These Gamma distribution parameters $\alpha, \beta$ can be estimated using the Expectation-Maximisation algorithm to find their optimal values [35].

## 4. Optimal Threshold using the EM Algorithm in Gamma Mixture Distribution

The application of the EM algorithm for the inference of probability distributions and their parameters is an effective and widely used tool [34–36]. To complete our deviation, we briefly describe the process to estimate $\alpha$ and $\beta$ in the Gamma distribution and to find the optimal threshold to convert an edge weight matrix into binary connections.

**E-Step:** We use $\alpha^{(n)}, \beta^{(n)}$ to represent the estimate of the parameters in the $n$th iteration of the EM algorithm. The expectation step calculates the expected value of the log-likelihood with respect to the hidden random variable $Y$,

$$\begin{aligned} &\mathscr{Q}\left(\alpha, \beta \mid \alpha^{(n)}, \beta^{(n)}, \mathbf{W}\right) = E_{Y \mid \alpha^{(n)}, \beta^{(n)}, \mathbf{W}}\{\mathscr{L}(\alpha, \beta \mid \mathbf{W}, \mathbf{Y})\} \\ &= \sum_{i=1}^{|E|} \sum_{j=0}^{j=1} p\left(Y_i = j \mid \omega_i, \alpha^{(n)}, \beta^{(n)}\right) \left(\log p\left(\omega_i \mid \alpha_j, \beta_j\right) + \log \pi_j\right) \end{aligned} \tag{4.1}$$

where $p(Y_i = j \mid \alpha, \beta)$ is the probability of $\omega_i$ to belong to the class $j$ at the $n$th iteration. This can be

derived from the Bayes rule as

$$p\left(Y_i = j \mid \omega_i, \alpha^{(n)}, \beta^{(n)}\right) = \frac{p\left(\omega_i \mid \alpha_j^{(n)}, \beta_j^{(n)}\right) p\left(Y_i = j \mid \alpha_j^{(n)}, \beta_j^{(n)}\right)}{G_W\left(\omega_i \mid \alpha_j^{(n)}, \beta_j^{(n)}\right)} \tag{4.2}$$

where the numerator is given Eq.(3.13),

$$\sum_{j=0}^{j=1} p\left(\omega_i \mid \alpha_j^{(n)}, \beta_j^{(n)}\right) p\left(Y_i = j \mid \alpha_j^{(n)}, \beta_j^{(n)}\right) = \sum_{j=0}^{j=1} \pi_j G_W(\omega_i \mid \alpha_j^{(n)}, \beta_j^{(n)}) \tag{4.3}$$

Then the maximization step can be decomposed as two independent steps for each term in Eq.(4.1). The first term depends on $\pi_j$ with the probability constraint that $\sum_{j=0}^{j=1} \pi_j = 1$. This is the likelihood term with $\pi_j$, and we use the method of Lagrange Multipliers to find the optimal solution [36].

**M-Step 1:** The derived function with the probability constraint in terms of the Lagrange multiplier $\lambda$ is given by

$$\mathscr{L}(\pi, \lambda) = \sum_{i=1}^{|E|} \sum_{j=0}^{j=1} p(Y_i = j \mid \omega_i, \alpha^{(n)}, \beta^{(n)}) \log \pi_j + \lambda \left( \sum_{j=0}^{i=1} \pi_j - 1 \right) \tag{4.4}$$

By calculating the derivatives with respect to $\pi_j$ and $\lambda$ respectively, and letting them equal to 0, we find the estimated value of $\hat{\pi}_j$ to be the solution of the Lagrange multiplier.

$$\hat{\pi}_j = \frac{1}{|E|} \sum_{i=1}^{|E|} p\left(Y_i = j \mid \alpha, \beta\right) \tag{4.5}$$

Next, we can maximize the log-likelihood in Eq.(4.1) which depends on $\alpha_j, \beta_j$.

**M-Step 2:** To estimate $\beta_j$ we first set the partial derivative of the log-likelihood for $\beta_j$ to zero

$$\frac{\partial}{\partial \beta_j} \left\{ \sum_{i=1}^{|E|} \sum_{j=0}^{j=1} p\left(Y_i = j \mid \alpha_j^{(n)}, \beta_j^{(n)}\right) \log p\left(\omega_i \mid \alpha_j, \beta_j\right) \right\} = 0 \tag{4.6}$$

where the log-likelihood of Gamma distribution in $p\left(\omega_i \mid \alpha_j, \beta_j\right)$ is

$$\log p\left(\omega_i \mid \alpha_j, \beta_j\right) = \left(\alpha_j - 1\right) \log \omega_i - \frac{\omega_i}{\beta_j} - \alpha_j \log\left(\beta_j\right) - \log\left(\Gamma\left(\alpha_j\right)\right) \tag{4.7}$$

This gives the optimal solution for parameter $\beta_j$

$$\beta_j = \frac{1}{\alpha_j} \frac{\sum_{i=1}^{|E|} \xi_{i,j} \omega_i}{\sum_{i=1}^{|E|} \xi_{i,j}} \tag{4.8}$$

where $\xi_{i,j} = p\left(Y_i = j \mid \alpha_j^{(n)}, \beta_j^{(n)}\right)$ is the short notation.

Similarly to estimate $\alpha_j$, we substitute Eq.(4.8) into Eq.(4.1) and set the resulting partial derivative to zero

$$\frac{\partial}{\partial \alpha_j}\left\{\sum_{i=0}^{|E|}\sum_{j=0}^{j=1}\xi_{i,j}\log p\left(\omega_i \mid \alpha_j, \frac{1}{\alpha_j}\frac{\sum_{i=1}^{|E|}\xi_{i,j}\omega_i}{\sum_{i=1}^{N}\xi_{i,j}}\right)\right\} = 0. \tag{4.9}$$

This gives the result,

$$\sum_{i=1}^{|E|}\xi_{i,j}\log(\omega_i) - \sum_{i=1}^{|E|}\xi_{i,j}\log\left(\frac{\sum_{k=1}^{N}\xi_{k,j}\omega_k}{\sum_{k=1}^{|E|}\xi_{k,j}}\right) + \sum_{i=1}^{|E|}\xi_{i,j}\log(\alpha_j) - \sum_{i=1}^{|E|}\xi_{i,j}\psi(\alpha_j) = 0, \tag{4.10}$$

where $\psi(x)$ is the *Digamma* function defined as $\Gamma'(x)/\Gamma(x)$.

Then, the corresponding solution for the optimal value of $\alpha_j$ is

$$\log(\alpha_j) - \psi(\alpha_j) = \log\left(\frac{\sum_{i}^{|E|}\xi_{i,j}\omega_i}{\sum_{i}^{|E|}\xi_{i,j}}\right) - \frac{\sum_{i}^{|E|}\xi_{i,j}\log\omega_i}{\sum_{i}^{|E|}\xi_{i,j}} \tag{4.11}$$

We can use an iterative numerical method to estimate the optional solution of $\hat{\alpha}_j$ in Eq.(4.11), and the estimate of $\hat{\beta}_j$ from Eq.(4.8). An alternative is to note that

$$\log(\alpha_j) - \psi(\alpha_j) = \log E[\omega_i] - E[\log\omega_i] \tag{4.12}$$

Making the approximation

$$\log\alpha_i - \frac{\Gamma'(\alpha_i)}{\Gamma(\alpha_i)} \simeq \frac{1}{\alpha_i}$$

which applies when $\alpha$ is small, we find that

$$\alpha_i \simeq \frac{1}{\log E[\omega_i] - E[\log\omega_i]} \tag{4.13}$$

as an approximate non-iterative solution to Eq.(4.11).

Using the estimated parameters from the EM algorithm, we can find the Gamma distribution mixture. This can generate the boundary solution for the threshold by solving the following equation

$$E[G_W(\omega^* \mid \alpha, \beta)] = G_W(\omega \mid \hat{\alpha}_1, \hat{\beta}_1) - G_W(\omega \mid \hat{\alpha}_2, \hat{\beta}_2) \tag{4.14}$$

This gives us the value of the threshold $\omega^*$ to convert the weighted network to an adjacency matrix.

## 5. Statistical Mechanics for Unweighted Network

For an unweighted network, each edge has a unit weight. The corresponding node degrees are analogous to the discrete energy states. The energy for each node is proportional to the nodal degree, that is

$$\omega_u = \omega k \tag{5.1}$$

where $\omega_u$ is the energy per node which is identical to the node weight; and $\omega = 1$ for an unweighted network. $k$ is the degree per node; and $k \in \mathbf{Z}$ which is a positive integer or zero and equal to the number of edges connecting to the node $u$. Thus, the occupation number of the energy states depends on the degree of the nodes connected by edges.

In the Boltzmann statistics, the nodes in the network are mapped to the particles in the thermal system. The probability distribution for individual nodes at the energy state can be given by the exponential function

$$P_u = \frac{1}{Z} e^{-\beta \omega_u} \tag{5.2}$$

where $Z$ is the partition function following the constrain of energy conservation

$$Z = \sum_{u=0}^{|V|} e^{-\beta \omega_u} \tag{5.3}$$

The average energy then can be derived from the corresponding partition function

$$\bar{U} = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} = -\frac{\partial \log Z}{\partial \beta} \tag{5.4}$$

This provides a framework to describe a network in the statistical ensemble with the thermal quantities, such as partition function and energy.

The derived temperature in Eq.(3.5) can also be extended to the networks, so that the thermodynamic partition function in Eq.(5.3) can be represented as a serial expansion

$$Z = \sum_{u=0}^{|V|} e^{-\beta \omega_u} = \frac{1 - e^{-|V|\beta\omega}}{1 - e^{-\beta\omega}} \approx \frac{1}{1 - e^{-\beta\omega}} \tag{5.5}$$

From Eq.(5.2), the probability of each node at a given energy state depends on the nodal degree

$$P(d_u = k) = \frac{1}{Z} e^{-\beta \omega_u} = \left(1 - e^{-\beta\omega}\right) e^{-\beta\omega k} \tag{5.6}$$

From Eq.(5.4), the corresponding energy related to the degree is given by

$$\bar{U} = \int_0^{\omega_T} \frac{|V|^2 \omega}{\pi} \cdot \frac{k^2}{e^{\beta\omega k} - 1} dk = \int_0^{\omega_T} P(\beta, k) dk \tag{5.7}$$

Therefore, the probability of each node given the degree $k$ and temperature $\beta$ is

$$P(\beta, k) = \frac{|V|^2 \omega}{\pi} \cdot \frac{k^2}{e^{\beta\omega k} - 1} = \frac{|V|^2}{2\pi|E|} \cdot \frac{k^2}{e^{\beta\omega k} - 1} \tag{5.8}$$

where $U = 2|E|\omega$. This describes the degree distribution in weighted networks not only relates to the nodal degree but also depends on the global parameter temperature as well.

## 6. Boundary and Temperature Limits

Because there is a limit on the total number of edges, we will now assume that the weight of nodes is possible up to a maximum boundary $\omega_T$. This is defined by

$$\int_0^{\omega_T} g(\omega) d\omega = 2|E| \tag{6.1}$$

which, using Eq.(3.9), implies that

$$\omega_T = \left(4\pi\frac{|E|}{|V|^2}\right)^{1/2}\omega \tag{6.2}$$

This allows us to rewrite Eq.(3.9) as

$$g(\omega)d\omega = \frac{4|E|\omega}{\omega_T^2}d\omega \tag{6.3}$$

Now we substitute Eq.(6.3) into Eq.(3.10) to get the logarithm of the partition function as

$$\log Z = -\frac{4|E|}{\omega_T^2}\int_0^{\omega_T}\omega\log\left[1-e^{-\beta w}\right]d\omega \tag{6.4}$$

According to Eq.(5.4), the average energy can be found that

$$\bar{U} = \frac{4|E|}{\omega_T^2}\int_0^{\omega_T}\frac{\omega^2}{e^{\beta w}-1}d\omega = \frac{4|E|}{\omega_T^2\beta^3}\int_0^{\frac{x_T}{\beta}}\frac{x^2}{e^x-1}dx \tag{6.5}$$

where $x = \beta\omega = \beta\omega k$. This equation is quite complicated and not obvious to see what the temperature dependence of average energy will be. This is because the exponential term is both degree and temperature dependent and the integral is degree dependent. But we can analyse the temperature limits for this equation.

***High-temperature Limit:*** At high temperature, $\beta \to 0$ and hence $e^x \to 1+x$. Hence, the average energy $\bar{U}$ behaves as

$$\bar{U} \to \frac{|V|^2}{\pi\omega^2\beta^3}\int_0^{\omega k} x dx = \frac{|V|^2}{2\pi}\cdot\frac{k^2}{\beta} \tag{6.6}$$

The corresponding nodal probability in Eq.(5.8) is

$$P(\beta,k) = \frac{|V|^2}{2\pi|E|}\cdot\frac{k}{\beta} \sim k\beta^{-1} \tag{6.7}$$

***Low-temperature Limit:*** At low temperature, $\beta \to \infty$ and hence $e^x \gg 1$. The average energy is given by

$$\bar{U} \to \frac{|V|^2}{\pi\omega^2\beta^3}\int_0^{\infty}\frac{x^2}{e^x}dx = \frac{|V|^2}{\pi\omega^2\beta^3}I_B(2) \tag{6.8}$$

where $I_B(2) = \zeta(3)\Gamma(3)$ is the Bose integral, which is the constant that $\zeta(3)$ is Riemann zeta function and $\Gamma(3)$ is the gamma function.

Then, the corresponding probability of node in Eq.(5.8) is

$$P(\beta,k) = \frac{C}{\omega^2\beta^3}\cdot\left(-\frac{1}{k_T^2}\right) \sim k_T^{-2}\beta^{-3} \tag{6.9}$$

where $C = |V|^2 I_B(2)/\pi$ is the constant, and $k_T = \omega_T/\omega$.

## 7. Experiments

### 7.1 *Datasets*

To analyse real-world data, we use both weighted and unweighted networks. Each kind of network contains four examples. For weighted networks, we study four specific examples, namely, a) airports in the United States (USAir)[37], b) Mammalia Hyena Network (MHN) [38], c) Condensed Matter Collaborations (CMC) [39], and d) Facebook-like Forum Network (FFN) [25]. For unweighted networks, the data comes from the complex networks in KONECT. These are the arXiv hep-ph network [40], the Facebook friendships network[41], the Google Orkut network [42] and Livemocha online language learning networks [43].

To provide a set of networks for our parameter clustering experiments, we use the data from synthetic and real-world networks. The synthetic groups of networks are generated from the typical network models containing Erdös-Rényi random graphs, Watts-Strogatz small-world networks [44], and Scale-free networks [2]. For the real-world data, we use the tumour mutation networks [45] and fMRI brain networks [46]. Each data contains a weighted network structure with different groups of patients.

#### 7.1.1 *Weighted Network Datasets*

***USAir:*** The United States air transport dataset contains the 500 airports with the actual air travelling flows among different urban regions [37]. The network represents the pair-wise airport connections with direct flights, where nodes are urban areas and the edges are air travel fluxes. The corresponding air transportation network comprises 332 nodes and 2126 edges. The edge weight corresponds to the frequency of flights between two airports.

***MHN:*** This Mammalia-hyena data comes from the Animal Social Network Repository(ASNR) [38]. This data recorded the interaction of mammalian hyenas in the real world within four months. The dataset contains 35 nodes and 521 edges. Spotted hyenas(Crocuta crocuta) are large mammalian carnivores, their societies are called 'clans'. The nodes in the data represent individual hyenas, and the edges are association patterns that were recorded based on the co-occurrence of each pair of individuals, during the period for which they were concurrently present in the clan. This animal social network was collected during periods of low prey abundance. The edge-weights represent the strength of social relations for pairs of hyenas, which is calculated by the ratio between the total association indices with all clan-mates to the sum of the number of other potential associations [38].

***CMC:*** This is the network of co-authorship who have submitted manuscripts to the e-Print Archive on the topic of condensed matter physics between 1995 and 1999. The network contains 16,726 nodes and 47,594 edges. Edge weights are estimated using Newman's method, i.e. $w_{ij} = \sum_k \delta_i^k \delta_j^k / (n_k - 1)$, where $n_k$ is the number of co-authors in the $k$th paper, and $\delta_i^k$ is 1 when the $i$th co-author appears the $k$th paper, and 0 otherwise.[39].

***FFN:*** The Facebook-like Forum Network contains users activity in the forum [25]. It comes from an online community with 899 users and 522 topics. Edge weights are assigned to the ties by considering the number of characters that a user posts to a topic. The edge weights are normalised between 0 and 1. This network contains 899 nodes and 33,720 edges [25].

#### 7.1.2 *Unweighted Network Datasets*

The arXiv hep-ph network is the collaboration graph of authors of scientific papers from arXiv's High Energy Physics-Phenomenology (hep-ph) section. An edge between two authors represents a common publication [40]. There are 28,093 nodes and 4,596,803 edges in the network. Facebook friendships network is the undirected network containing the friendship

data of Facebook users. A node represents a user and an edge represents a friendship between two users [41]. There are 63,731 nodes and 817,035 edges in the network. The Google Orkut network is the social network of Orkut users and their connections. There are 3,072,441 nodes and 117,185,083 edges in the network [42]. The Livemocha dataset consists of social networks which describe the friendships in the world's largest online language community [43]. There are 104,103 nodes and 2,193,083 edges in the network.

#### 7.1.3 *Network Group Datasets*

***Synthetic Networks:*** The synthetic networks include three groups, i.e., Erdös-Rényi random graphs, Watts-Strogatz small-world networks [44], and Scale-free networks [2]. Each network contains 1,000 nodes, and there are 500 networks in each group. For the random graph, the probability of connection between two nodes is 0.5. For the small-world model, the mean nodal degree is 50 with a rewiring probability 0.15. For the scale-free model, the probability of adding a new node connected to an existing node is 0.7, and the probability of adding an edge between two existing nodes is 0.2. The edge weights are computed according to the node degree as $w = d_u d_v / E$, where $d_u$ and $d_v$ are nodal degrees between two ends of the edge.

***Tumour Mutation Networks:*** The tumour mutation dataset contains networks representing gene mutation patterns for three major cancers taken from Cancer Genome Atlas (TCGA). These are a) ovarian cancer, b) uterine cancer and c) lung adenocarcinoma [47]. There are 356 subjects with mutations in 9,850 genes in the ovarian cancer cohort, 248 subjects with mutations in 17,968 genes in the uterine endometrial cancer cohort and 381 subjects with mutations in 15,967 genes in the lung adenocarcinoma cohort [45]. Each subject is characterized by a sequence of gene indicators. The mutation networks were mapped onto gene interactions by aggregating information from several pathways and interaction databases, describing physical protein-protein interactions (PPIs) and functional relationships between genes in both regulatory, signalling and metabolic pathways [48, 49]. The edge weights are the strengths of interactions between different genes.

***Brain Networks:*** The brain network data comes from fMRI images in the ADNI database. These record the Blood-Oxygenation-Level-Dependent (BOLD) signals for different anatomical brain regions [46]. The Anatomical Automatic Labeling atlas (AAL) is used as a template to separate the brain into 90 regions of interest (ROIs) [50]. Weighted networks are constructed by computing the cross-correlation coefficients between the BOLD signals between each pair of ROIs. There are two categories of patients in the Alzheimer's disease study. One group has 105 subjects with fully developed Alzheimer's disease (AD), and the second group has 193 normal healthy control subjects (NC).

### 7.2 *Experimental Evaluation*

#### 7.2.1 *Weighted Network Analysis*

We first provide a numerical analysis of the density distribution of edge weights in Eq.(3.12). This shows how the probability density function varies with the edge weight $\omega$ and inverse temperature $\beta$.

Fig.1 plots the three-dimensional variation in the edge-weight probability with edge weight and temperature. The overall probability increases with the value of edge weights. This is especially obvious in the high-temperature region (low value of $\beta$). In the meanwhile, the probability exponentially decays as the temperature decreases (inverse temperature $\beta$ increases).

An important feature is that there is a nontrivial phase transition that occurs in the low-temperature region. We show this in Fig.2 by fixing the value of $\beta = 1, 1.5, 2$. As shown in Fig.2, the probability of edge weight increases up to a maximum and then decreases in the low-temperature region. When the

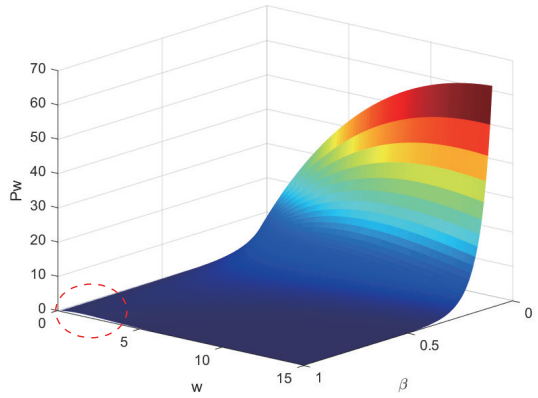


FIG. 1. Three dimensional plot of the probability density function for edge weights. There is a phase transition in the low temperature region (red circle).

value of inverse temperature increases, this peak corresponding to the phase transition shifts towards zero. Since the parameter of temperature is proportional to the edge density in the network, this shows that the phase transition is more likely to happen in sparsely connected networks.

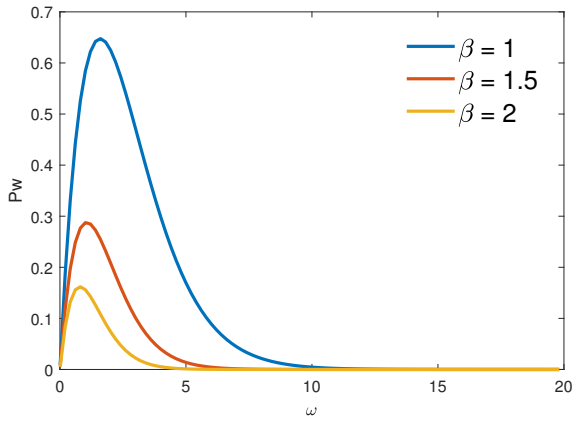


FIG. 2. The probability function changes with the edge weight in the lwo temperature region. The parameter of inverse temperature is fixed at $\beta = 1, 1.5, 2$.

We now turn our attention to real-world datasets. To explore the nontrivial phase transition described above, we plot the histogram of edge weights for four types of real-world networks. These networks have high values of $\beta$ which means that the corresponding temperatures are low. Fig.3(a) shows the distribution of normalized edge weights for the USAir, MHN and FFN datasets. We observe a similar transition pattern as shown in Fig.2. This verifies that small edge weights exist with a higher probability in low-temperature networks. A more extreme example is shown in Fig.3(b) for the CMC network. Here we observe a large value of $\beta$ corresponding to an extremely low temperature, and where there is

a large fraction of edge weights having very small values.

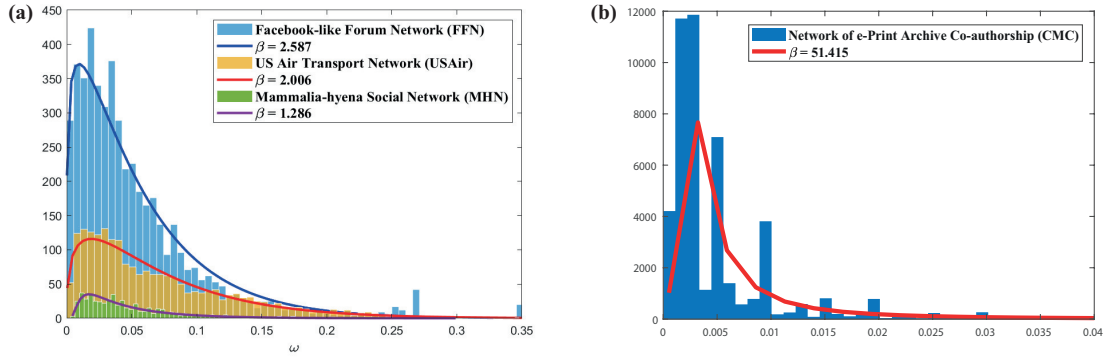


FIG. 3. The histogram of edge weights in the real-world networks with high value of $\beta$. (a) United States air transport network ($\beta$=2.006), Mammalia-hyena animal network ($\beta$=1.286), Facebook-like forum network ($\beta$=2.587). (b) Extreme low temperature case: co-authorship network in condensed matter physics of arXiv e-prints ($\beta$=51.415).

Furthermore, we examine the histogram of edge weight distribution with the Gamma distribution functions and represent this function as a mixture by estimating the parameters using the EM algorithms with two Gamma mixture components. The overlap boundary of these two Gamma distributions will be used to set the optimal threshold to infer binary connection indicators from edge weights.

Fig.4 shows the edge-weight distributions for the four different real-world networks. These networks come from the complex network datasets, which are the United States air transport network, Mammalia-hyena animal social network, co-authorship in e-Print in condensed matter physics of arXiv e-prints, and Facebook-like forum network. In Fig.4, the black dots indicate the link strength is the original weight distribution. The blue curve is the generalised Gamma function fitted to the edge-weight distribution. The red curve is the Gamma mixture distribution with two different Gamma functions. The decomposition of each distribution can be applied to the EM algorithm to estimate the corresponding parameters. This is shown by the green curves. We observe that the edge-weight distribution in real-world networks fits well with our derived probability density function. This distribution can be approximated and further decompose the combination of Gamma distribution functions into a mixture.

Finally, we set the threshold by decomposing the mixture Gamma distribution into two mixing components. This can be calculated using Eq.(4.11) to set the threshold for the edge weights. We take the CMC dataset as an example to compare the structures resulting in weighted and unweighted representations of the networks. Fig.5 shows the network structure before and after setting the threshold.It is clear to observe that the proposed threshold can keep the main structure of the weighted network, which can be applied to reduce the redundant information and to figure out the backbone of the weighted network.

### 7.2.2 *Unweighted Network Analysis*

For unweighted networks, we conduct the numerical analysis on the node probability in Eq.(5.6). Fig.6 shows how the node probability changes with the degree $k$ and inverse temperature $\beta$, respectively. In Fig.6(a), there is a phase transition for the node probability as the node degree varies. As the inverse temperature $\beta$ increases the peak value of the node probability moves towards zero. In Fig.6(b), the node probability decays exponentially with the inverse temperature. The larger the value of the nodal degree, the faster the exponential decay.

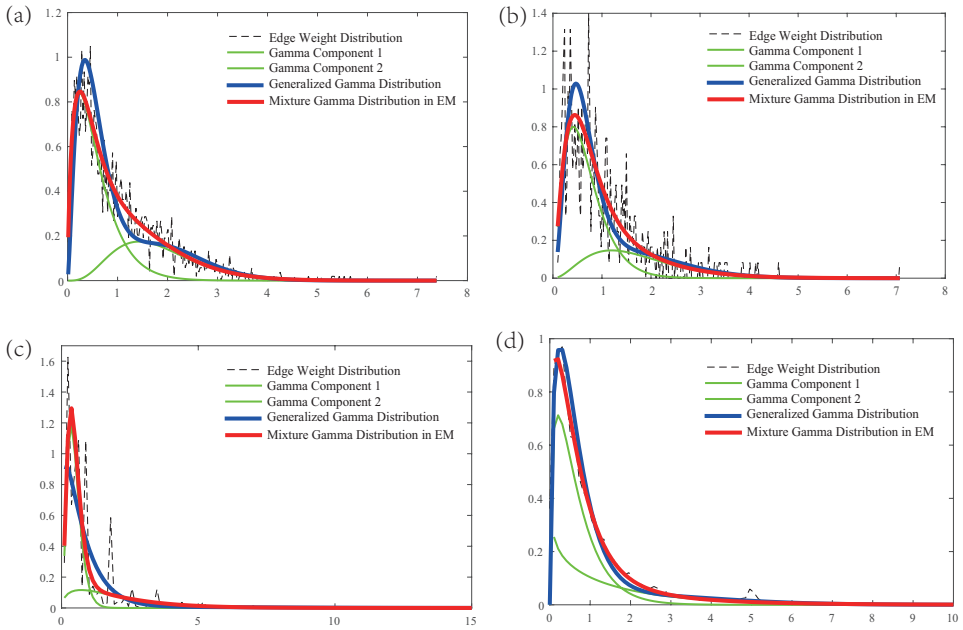


FIG. 4. Applying the mixture Gamma functions to fit the edge weight distribution on the real-world networks. The black line is the original weight distribution. The blue curve is the generalized Gamma distribution to fit the original weights. The red curve is the mixture of Gamma distribution with the EM algorithm. Green curves are two decomposed components. The cross point of two green line components generates the value of the threshold to convert weighted networks to binary connections. (a) The United States air transport network; (b) Mammalia-hyena animal network; (c) co-authorship network in condensed matter physics of arXiv e-prints; (d) Facebook-like forum network.

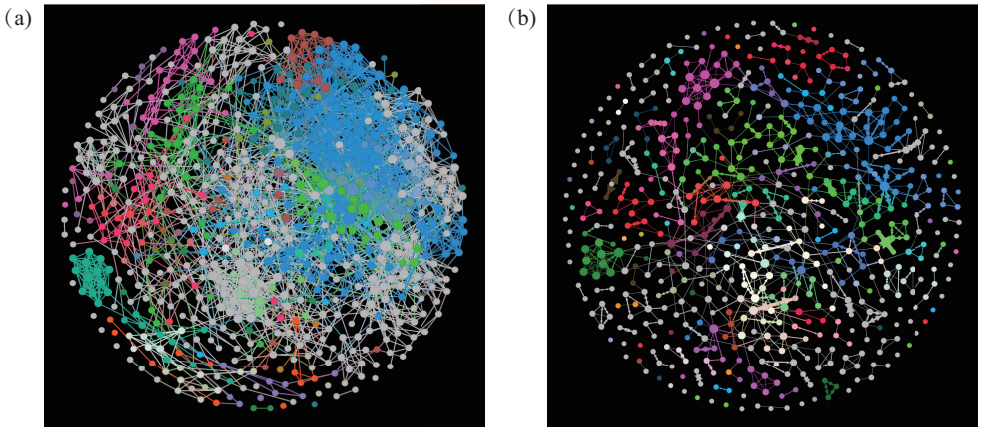


FIG. 5. The visualization of co-authorship network before and after thresholding. Edge weights are normalised between 0 and 1, and the threshold value is 0.5121. (a) Original weighted network. (b) Binary connections after thresholding.

We now analyse the energy in Eq.(5.7) and explore how it varies with node degree and temperature. The expression in Eq.(5.7) is quite complicated and the functional dependence on these two quantities is

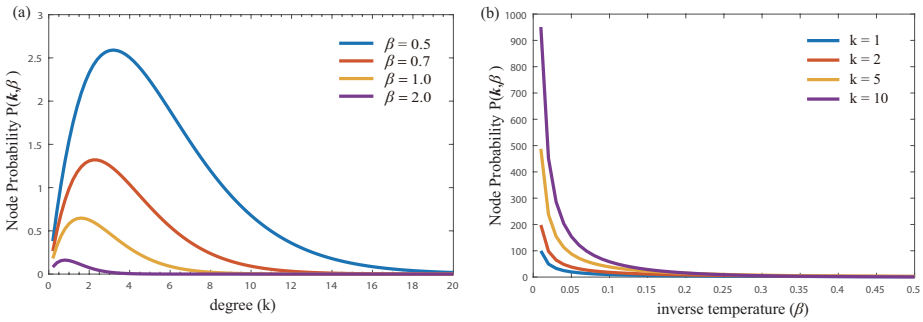


FIG. 6. The node probability changes with the degree $k$ and inverse temperature $\beta$ in Eq.(5.6). (a) node probability with degree; (b) node probability with inverse temperature

not obvious by inspection. The reason for this is that the exponential term is both degree and temperature dependent and the integral is also degree dependent. In Fig.7, we show the full degree of dependence for the energy. The energy increases with the degree while reaching a constant value when the node degree becomes large. The energy decreases rapidly as the inverse temperature $\beta$ increases.

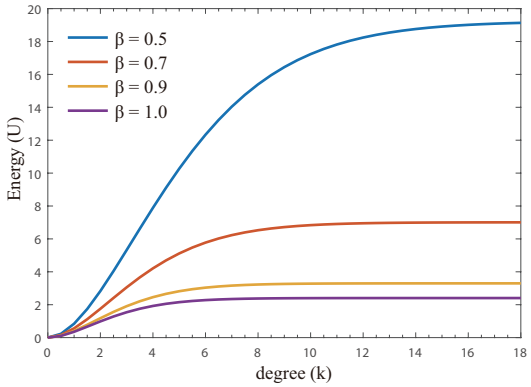


FIG. 7. The network energy changes with the degree according to Eq.(5.7)

Finally, we turn our attention to the real-world datasets. We investigate the node probability in Eq.(5.6)for the complex network dataset. Fig.8 shows the degree distributions for four different complex networks. The blue curves are the actual degree distributions and the red curves are the predictions of our model. The four networks come from the KONECT which are the arXiv hep-th networks, the Facebook network, the Google Orkut user network, and Livemocha social networks. It is clear that instead of following a power law the degree distribution is better fitted by the predictions of Eq.(5.6) for low values of node degree. At the high-degree end, the distribution more closely follows the power law.

Our derived expression for the node degree distribution can be used to fit the degree distributions of real complex networks. The corresponding energy and temperature are associated with the network structure.

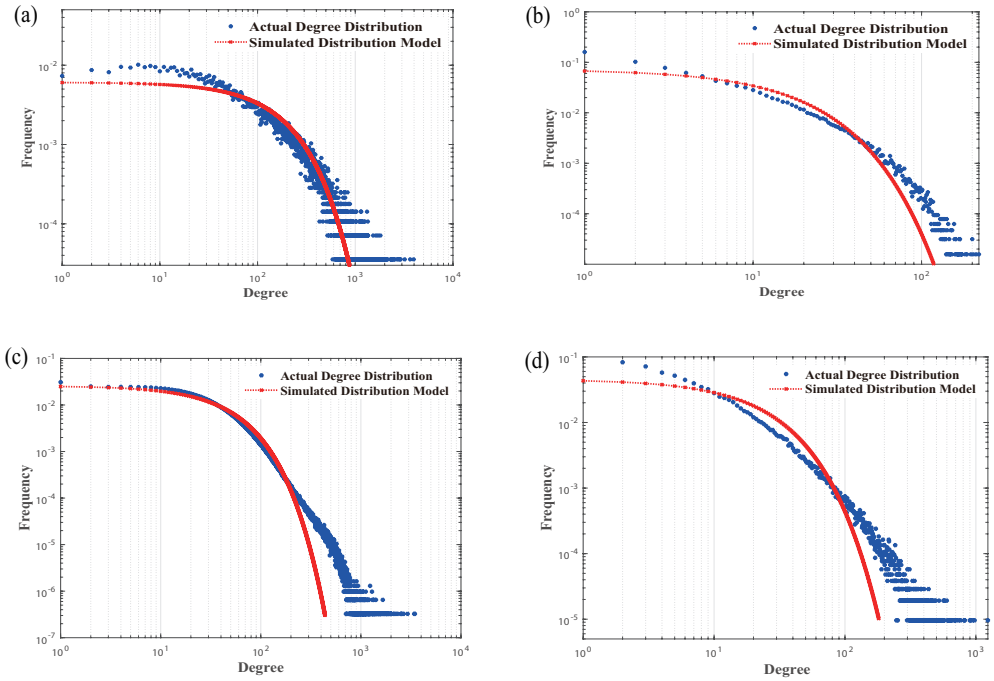


FIG. 8. Degree distribution of real-world networks. The blue curves are actual degree distributions and the red curves are the simulation from Eq.(5.6). (a) ArXiv's High Energy Physics–Phenomenology (hep-ph) containing 28,093 nodes and 4,596,803 edges. (b) Facebook networks with 63,731 nodes and 817,035 edges. (c) the collection of social network in Google Orkut users, consisting of 3,072,441 nodes and 117,185,083 edges. (d) Livemocha social network with 104,103 nodes and 2,193,083 edges.

## 7.3 *Network Group Analysis*

### 7.3.1 *Synthetic Network Evaluation*

Here, we investigate whether the estimated parameters $\alpha$ and $\beta$ can be used to cluster different types of networks. We generate synthetic networks in three groups, i.e., Erdös-Rényi random graphs, Watts-Strogatz small-world networks [44], and Scale-free networks [2]. Each network contains 1,000 nodes, and there are 500 networks in each group.

After applying the EM algorithm to estimate edge weight distribution in the Gamma mixture model, we evaluate whether the two pairs of parameters $\alpha$ and $\beta$ for the different mixing components can be used to group different network structures. Fig.9 shows the three dimensional scatter plot of $\alpha_1$, $\alpha_2$ and $\beta_1$ for three groups. Since the scale-free network follows the power-law distribution, the shape parameter $\alpha$ plays the dominant role in distinguishing this structure. The small-world network and the Erdos-Renyi random graph have a similar distribution when the rewiring probability approaches unity. The estimated parameters can still be used to cluster these different networks into two groups.

### 7.3.2 *Tumour Mutation Network*

Next, we turn our attention to the gene mutation networks. These contain three different groups of gene interaction networks in cancer, i.e., lung adenocarcinoma, ovarian cancer and uterine endometrial. In Fig.10, we visualise the three-dimensional space of the estimated parameters in the Gamma mixture

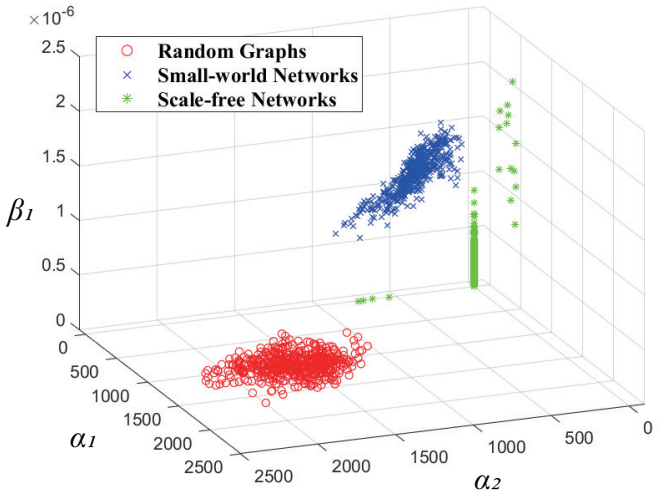


FIG. 9. Three dimentional scatter plot of $\alpha_1$, $\alpha_2$ and $\beta_1$ for different network models. Erdös-Rényi random graphs in red circle, Small-world networks in blue cross, and Scale-free networks in green star.

model. The different coloured points represent the three different types of cancers for the subjects.

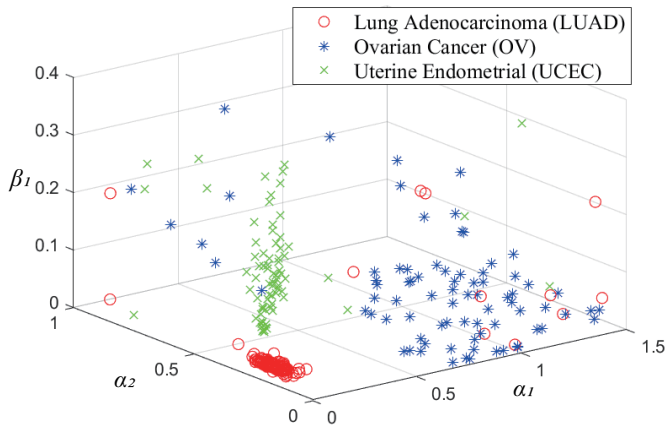


FIG. 10. Three dimentional scatter plot of $\alpha_1$, $\alpha_2$ and $\beta_1$ for tumour mutation networks (lung adenocarcinoma in red circle, ovarian cancer in blue star, and uterine endometrial in green cross).

The scatter plot reveals clusters corresponding to the three different classes of tumour mutation networks. The clusters exhibit a compact statistical structure, with little cross-class contamination. Moreover, the different groups of cancer networks are well separated in the space of parameters of $\alpha_1$, $\alpha_2$ and $\beta_1$. This is especially so for the lung adenocarcinoma and uterine endometrial with two

distinguished high-density clusters. This further illustrates the effectiveness of the estimated parameters in distinguishing different classes of real-world networks.

### 7.3.3 *fMRI Brain Networks*

To analyse the distribution of the edge weight parameters, we plot the histogram of $\alpha$ and $\beta$ for fMRI brain networks. There are two groups in the sample studied, i.e. those with Alzheimer's disease and the normal control sample. As shown in Fig.11(a) and Fig.11(b), although the separate distribution of $\alpha$ and $\beta$ are overlapped for the two groups of patients, there are two-well separated parameter clusters. This is because we treat the edge weight distribution as a two-component Gamma mixture distribution. Each set of parameters $\alpha$ and $\beta$ is estimated from one of the Gamma components.

We also observe the distribution of threshold between the two groups of subjects. In Fig.11(c), the AD subjects have a narrower and more concentrated spread of thresholds when compared to the normal healthy sample.

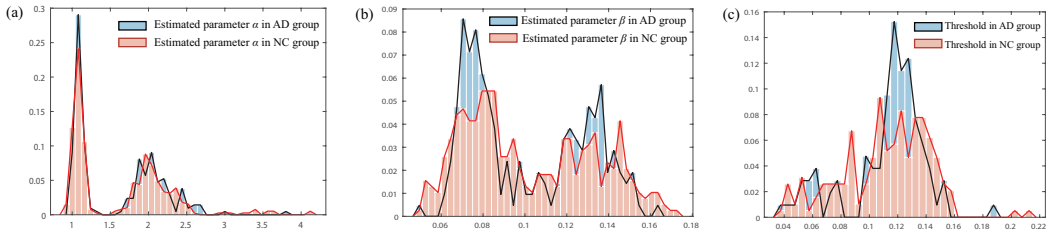


FIG. 11. The distribution of parameters $\alpha$ and $\beta$ in the mixed Gamma functions between two groups of patients. The blue areas are the histogram of weights in AD group and the red areas are the histogram of edge weights in normal healthy group. (a) The distribution of parameter $\alpha$. (b) The distribution of parameter $\beta$. (c) The distribution of the value of threshold.

We can also investigate the relationship between the two parameters $\alpha$ and $\beta$ for two groups of Alzheimer's subjects (AD and NC). In Fig.12(a) and Fig.12(b), the scatter plots show that both the iterative numerical method and the non-iterative approximation for $\alpha$ present a similar linear correlation with $\beta$, especially when the value is small. This is because we decompose the distribution of weights into a mixture of two Gamma functions. The distribution of small values of weights presents a very similar pattern and the shape of this distribution depends on the parameter $\alpha$.

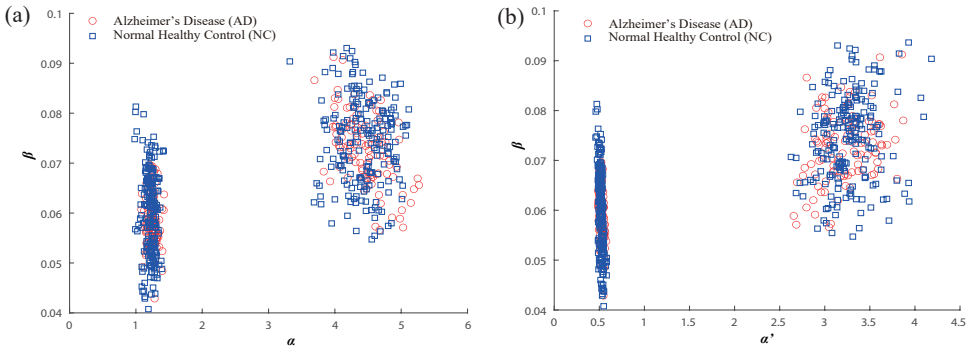


FIG. 12. The scatter plot of parameters $\alpha$ and $\beta$ in the mixed Gamma functions between two groups of subjects. (a) The distribution of iterative numerical estimation of $\alpha$ with $\beta$. (b) The distribution of non-iterative approximation of $\alpha$ with $\beta$.

Therefore, we conclude that the estimated parameters for the edge weight Gamma mixture distribution can be used as features to distinguish different types of network structures.

### 7.4 *Discussion*

To demonstrate the advantages of our proposed method, we compare the corresponding results to both typical and state-of-the-art thresholding methods. Two of these methods are global threshold selection using a fixed value or maintaining the edge density. An alternative method is the Disparity Filter which applies a localized threshold [51]. The final method studied is structural inference to locate true connections using the edge weights [14]. These methods serve as baselines for comparison.

Here, we compare our proposed method to different baselines. We measure four commonly used statistical properties in complex networks, i.e., average degree, clustering coefficients [52], average path length and communicability [53] for the real-world weighted networks. This reveals the different effects of these models after thresholding the structure of the edge weight distributions. In Table1, we present a comparison of results for the USAir, MHN, CMC and FFN networks.

Thresholding with a fixed value or edge density has identical effects on corresponding networks. Although these are simple and intuitive approaches, they may lose important network structures. This is illustrated in Table1 where both methods fail to preserve all five structural properties for the reconstructed networks. The measurements of average degree and clustering coefficients reveal this performance. These demerits are obvious in the datasets of USAir, CMC and FFN, especially when the size of networks becomes large.

Local thresholding methods such as the Disparity Filter [51], provide potential solutions to the problem of loss of fat-tailed weight distributions of edge weights. The method computes the value of the threshold at the local level of each node, rather than over the entire network. In the table for the USAir dataset, the Disparity Filter preserves the backbone of the networks by slightly improving the clustering coefficients and average path length.

However, simple thresholding can corrupt the network's structural properties. Specifically, the network connectivity may become sparse and begin to fragment with an unreasonable choice of threshold. Statistical inference estimates the community structure on uncertain networks [14]. This improves structural inference from the weighted network to give more reliable edges. In Table1, when the networks become large, the structural properties from uncertain inference improve the performance in terms of average path length and communicability. This is most obvious for the CMC and FFN networks.

When compared to the alternative method, our proposed approach improves the statistical structural properties, especially communicability. As well as outperforming in terms of the average degree and clustering coefficients for the MHN and FFN networks, we also preserve better network structure in terms of communicability. This is best observed in the USAir, MHN and FFN datasets. Even though uncertain inference performs well on large-scale complex networks, such as CMC, our model exhibits better performance in each of the structural properties.

Our novel approach based on the mixture Gamma model fits accounts well for the empirically observed distribution of edge weights. Compared to alternative thresholding methods, our approach provides an optimal way to filter out redundancy and maintain the network backbone structure with better performance in terms of describing the observed statistical properties.

Table 1. Comparsion across other thresholding methods in real-wrorld weighted networks with different structural properties

| Network | Nodes | Threshold Model | Edges | Ave. Degree | Clust. Coeff. | Ave. Path Length | Communicability |
|---|---|---|---|---|---|---|---|
| U.S. Air Transport Network (USAir) | 332 | Original | 2126 | 12.807 | 0.625 | 1.738 | $8.081 \times 10^{17}$ |
| | | Fixed Value | 853 | 5.139 | 0.163 | 1.442 | $2.103 \times 10^{11}$ |
| | | Density | 853 | 5.139 | 0.163 | 1.442 | $2.103 \times 10^{11}$ |
| | | Disparity Filter | 863 | 5.103 | **0.205** | **2.150** | $1.957 \times 10^{10}$ |
| | | Uncertain Inference | 848 | 5.111 | 0.163 | 1.442 | $1.968 \times 10^{11}$ |
| | | **Our Model** | 875 | **5.271** | 0.173 | 1.430 | $\mathbf{3.283 \times 10^{11}}$ |
| Mammalia-hyena Social Network(MHN) | 35 | Original | 521 | 29.771 | 0.922 | 0.124 | $1.736 \times 10^{13}$ |
| | | Fixed Value | 210 | 12.000 | 0.558 | **0.735** | $2.760 \times 10^{6}$ |
| | | Density | 210 | 12.000 | 0.558 | **0.735** | $2.760 \times 10^{6}$ |
| | | Disparity Filter | 209 | 11.943 | 0.422 | 0.671 | $1.193 \times 10^{6}$ |
| | | Uncertain Inference | 259 | 14.809 | 0.623 | 0.612 | $3.851 \times 10^{7}$ |
| | | **Our Model** | 270 | **15.429** | **0.660** | 0.599 | $\mathbf{6.025 \times 10^{7}}$ |
| Condensed Matter Collaborations(CMC) | 16,726 | Original | 47594 | 5.853 | 0.638 | 5.627 | $7.437 \times 10^{10}$ |
| | | Fixed Value | 25890 | 3.184 | 0.467 | **6.395** | $7.131 \times 10^{5}$ |
| | | Density | 25890 | 3.184 | 0.467 | **6.395** | $7.131 \times 10^{5}$ |
| | | Disparity Filter | 26239 | 3.119 | 0.295 | 6.161 | $2.095 \times 10^{5}$ |
| | | Uncertain Inference | 26726 | **3.287** | **0.530** | 6.122 | $\mathbf{1.511 \times 10^{6}}$ |
| | | **Our Model** | 26649 | 3.277 | 0.475 | 6.356 | $1.456 \times 10^{6}$ |
| Facebook-like Forum Network(FFN) | 899 | Original | 7046 | 15.675 | 0.076 | 1.832 | $1.699 \times 10^{14}$ |
| | | Fixed Value | 1717 | 3.820 | 0.009 | 2.947 | $7.641 \times 10^{4}$ |
| | | Density | 1717 | 3.820 | 0.009 | 2.947 | $7.641 \times 10^{4}$ |
| | | Disparity Filter | 1820 | 4.017 | 0.011 | 2.610 | $5.603 \times 10^{4}$ |
| | | Uncertain Inference | 1486 | 3.830 | 0.013 | **3.104** | $1.905 \times 10^{4}$ |
| | | **Our Model** | 1946 | **4.329** | **0.026** | 2.818 | $\mathbf{1.567 \times 10^{5}}$ |

## 8. Conclusion

In this paper, we describe a new method to infer binary edge indicators from the distribution of edge weights via the setting of an optimal threshold. To this end, we make use of the statistical mechanical model to describe the probability distribution function for edge weights. Using this idea, we apply the network structure to define a temperature parameter. This provides the physical interpretation of temperature in terms of the number of nodes and edges. This is proportional to the edge density. We can further apply the statistical mechanical approach to derive further derive additional physical parameters of the network.

Using an analogy in which the nodes of the network are particles in the solid model, edge weights are analogous to the microstate energies in a two-dimensional solid lattice. This allows us to derive a probability density function for edge weights which depends on the values of the edge weights and the global temperature parameter which is related to the configuration of nodes and edges. The corresponding distribution function is represented in closed form as a Gamma distribution. We represent the distribution function by a mixture of Gamma functions, and extract the mixing proportions and parameters of the mixture components using the Expectation-Maximization algorithm. Minimizing the overlap between two Gamma distributions provides the optimal value for inferring a set of binary edge indicators from the distribution of edge weights.

Furthermore, the edge weights can be regarded as being uniform in unweighted networks. This gives an exponential expression for the probability density function for nodes, which is identical to the degree distribution. It depends on both the edge weights and the global parameter of temperature related to the configuration of nodes and edges. The nodal probability density function together with the cumulative distribution function for the energy reveals a phase transition for both the degree and temperature dependence.

We conduct numerical experiments that demonstrate the existence of this phase transition in network sparsity that occurs in the low-temperature regime. Experimental results on real-world weighted

network datasets reveal the distribution of edge weights fits well with our derived function with a two Gamma function mixture model. The threshold can be further used to filter out redundancy and maintain the network backbone structure.

## 9. Acknowledgement

This work is sponsored by Shanghai Pujiang Program (No.21PJ1404200), Overseas Visiting Fellow Program (No.22WZ2505000).

## REFERENCES


1. Jianjia Wang, Hui Wu, and Edwin R Hancock. Thermal characterisation of unweighted and weighted networks. *IEEE 25th International Conference on Pattern Recognition (ICPR)*, pages 1641–1648, 2021.
2. Anna D Broido and Aaron Clauset. Scale-free networks are rare. *Nature Communications*, 10(1):1–10, 2019.
3. Charles Murphy, Antoine Allard, Edward Laurence, Guillaume St-Onge, and Louis J Dubé. Geometric evolution of complex networks with degree correlations. *Physical Review E*, 97(3):032309, 2018.
4. Giulio Cimini, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and Guido Caldarelli. The statistical physics of real-world networks. *Nature Reviews Physics*, 1(1):58–71, 2019.
5. Xiaomin Wang, Fei Ma, and Bing Yao. Dense networks with mixture degree distribution. *Frontiers in Physics*, 9:111, 2021.
6. Y. Gao, Z. Zhang, H. Lin, X. Zhao, S. Du, and C. Zou. Hypergraph learning: Methods and practices. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 44(05):2548–2566, 2022.
7. Illés Farkas, Dániel Ábel, Gergely Palla, and Tamás Vicsek. Weighted network modules. *New Journal of Physics*, 9(6):180, 2007.
8. Min Li, Jian-Xin Wang, Huan Wang, and Yi Pan. Identification of essential proteins from weighted protein–protein interaction networks. *Journal of Bioinformatics and Computational Biology*, 11(03):1341002, 2013.
9. Jianjia Wang, Jiayu Huo, and Lichi Zhang. Thermodynamic edge entropy in alzheimer's disease. *Pattern Recognition Letters*, 125:570–575, 2019.
10. Jean-Gabriel Young, George T Cantwell, and MEJ Newman. Bayesian inference of network structure from unreliable data. *Journal of Complex Networks*, 8(6):cnaa046, 2020.
11. Leto Peel, Tiago P Peixoto, and Manlio De Domenico. Statistical inference links data and theory in network science. *Nature Communications*, 13(1):6794, 2022.
12. George T Cantwell, Yanchen Liu, Benjamin F Maier, Alice C Schwarze, Carlos A Serván, Jordan Snyder, and Guillaume St-Onge. Thresholding normally distributed data creates complex networks. *Physical Review E*, 101(6):062302, 2020.
13. Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1), 2007.
14. Travis Martin, Brian Ball, and M. E. J. Newman. Structural inference for uncertain networks. *Phys. Rev. E*, 93:012306, Jan 2016.
15. Pranay Sharma, Donald J Bucci, Swastik K Brahma, and Pramod K Varshney. Communication network topology inference via transfer entropy. *IEEE Transactions on Network Science and Engineering*, 7(1):562–575, 2019.
16. Canh Hao Nguyen and Hiroshi Mamitsuka. Learning on hypergraphs with sparsity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8):2710–2722, 2020.
17. Jianjia Wang, Xin Zhao, Chong Wu, and Edwin R Hancock. Inferring edges from weights in the debye model. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 3845–3850. IEEE, 2022.
18. Edmund T Rolls, Chu-Chung Huang, Ching-Po Lin, Jianfeng Feng, and Marc Joliot. Automated anatomical labelling atlas 3. *NeuroImage*, 206:116189, 2020.
19. Tiago Simas, Rion Brattig Correia, and Luis M Rocha. The distance backbone of complex networks. *Journal of Complex Networks*, 9(6), 10 2021.

**20.** Guido Previde Massara, Tiziana Di Matteo, and Tomaso Aste. Network filtering for big data: Triangulated maximally filtered graph. *Journal of complex Networks*, 5(2):161–178, 2016.

**21.** Kamal Berahmand, Mehrnoush Mohammadi, Farid Saberi-Movahed, Yuefeng Li, and Yue Xu. Graph regularized nonnegative matrix factorization for community detection in attributed networks. *IEEE Transactions on Network Science and Engineering*, 2022.

**22.** Mark Drakesmith, Karen Caeyenberghs, A Dutt, G Lewis, AS David, and Derek K Jones. Overcoming the effects of false positives and threshold bias in graph theoretical analyses of neuroimaging data. *Neuroimage*, 118:313–333, 2015.

**23.** Xin Luo, Zhigang Liu, Mingsheng Shang, Jungang Lou, and MengChu Zhou. Highly-accurate community detection via pointwise mutual information-incorporated symmetric non-negative matrix factorization. *IEEE Transactions on Network Science and Engineering*, 8(1):463–476, 2020.

**24.** Mark EJ Newman. Spectral methods for community detection and graph partitioning. *Physical Review E*, 88(4):042822, 2013.

**25.** Tore Opsahl. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, 35(2):159–167, 2013.

**26.** Shiping Wen, Huaqiang Wei, Zheng Yan, Zhenyuan Guo, Yin Yang, Tingwen Huang, and Yiran Chen. Memristor based design of sparse compact convolutional neural network. *IEEE Transactions on Network Science and Engineering*, 7(3):1431–1440, 2019.

**27.** Benjamin D Haeffele and René Vidal. Structured low-rank matrix factorization: Global optimality, algorithms, and applications. *IEEE Trans. on pattern analysis and machine intelligence*, 42(6):1468–1482, 2019.

**28.** David M Walker and Débora C Corrêa. Network of compression networks to extract useful information from multivariate time series. *Journal of Complex Networks*, 11(3), 05 2023.

**29.** Christopher Aicher, Abigail Z Jacobs, and Aaron Clauset. Learning latent block structure in weighted networks. *Journal of Complex Networks*, 3(2):221–248, 2015.

**30.** Yuan Sun, Xiaodong Li, and Andreas Ernst. Using statistical measures and machine learning for graph reduction to solve maximum weight clique problems. *IEEE Trans. on pattern analysis and machine intelligence*, 43(5):1746–1760, 2019.

**31.** Sikun Yang and Heinz Koeppl. A poisson gamma probabilistic model for latent node-group memberships in dynamic networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

**32.** Valérie Poulin and François Théberge. Comparing graph clusterings: Set partition measures vs. graph-aware measures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 43(6):2127–2132, 2020.

**33.** Jianjia Wang, Richard C Wilson, and Edwin R Hancock. Spin statistics, partition functions and network entropy. *Journal of Complex Networks*, 5(6):858–883, 07 2017.

**34.** Derek S Young, Xi Chen, Dilrukshi C Hewage, and Ricardo Nilo-Poyanco. Finite mixture-of-gamma distributions: estimation, inference, and model-based clustering. *Advances in Data Analysis and Classification*, 13:1053–1082, 2019.

**35.** Sivaraman Balakrishnan, Martin J Wainwright, and Bin Yu. Statistical guarantees for the em algorithm: From population to sample-based analysis. *The Annals of Statistics*, 45(1):77–120, 2017.

**36.** Gonzalo Vegas-Sanchez-Ferrero, Marcos Martin-Fernandez, and Joao Miguel-Sanches. *A Gamma Mixture Model for IVUS Imaging*. Multi-Modality Atherosclerosis Imaging and Diagnosis, 2014.

**37.** Vittoria Colizza, Romualdo Pastor-Satorras, and Alessandro Vespignani. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, 3(4):276–282, 2007.

**38.** Kay E Holekamp, Jennifer E Smith, Christopher C Strelioff, Russell C Van Horn, and Heather E Watts. Society, demography and genetic structure in the spotted hyena. *Molecular Ecology*, 21(3):613–632, 2012.

**39.** Mark EJ Newman. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, 98(2):404–409, 2001.

**40.** Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge Discovery in Data Mining*, pages 177–187, 2005.

**41.** Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P Gummadi. On the evolution of user interac-

tion in facebook. *Proceedings of the 2nd ACM workshop on Online social networks*, pages 37–42, 2009.

**42.** Jaewon Yang and Jure Leskovec. Defining and evaluating network communities based on ground-truth. *Knowledge and Information Systems*, 42(1):181–213, 2015.

**43.** R. Zafarani and H. Liu. Social computing data repository at ASU, 2009.

**44.** Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

**45.** Matan Hofree, John P Shen, Hannah Carter, Andrew Gross, and Trey Ideker. Network-based stratification of tumor mutations. *Nature Methods*, 10:1108–1115, 2013.

**46.** Ronald Carl Petersen, PS Aisen, et al. Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology*, 74(3):201–209, 2010.

**47.** Junjun Zhang, Rosita Bajari, Dusan Andric, Francois Gerthoffert, Alexandru Lepsa, Hardeep Nahal-Bose, Lincoln D Stein, and Vincent Ferretti. The international cancer genome consortium data portal. *Nature Biotechnology*, 37(4):367–369, 2019.

**48.** Cerami Ethan, Gross Benjamin, Demir Emek, Rodchenkov Igor, Babur Ozgun, Anwar Nadia, Schultz Nikola, Bader Gary, and Sander Chris. Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res*, 39:D695–D690, 2011.

**49.** Insuk Lee, U Martin Blom, Peggy I Wang, Jung Eun Shim, and Edward M Marcotte. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Research*, 21(7):1109–1121, 2011.

**50.** Rolf A Heckemann, Joseph V Hajnal, Paul Aljabar, Daniel Rueckert, and Alexander Hammers. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, 2006.

**51.** M Ángeles Serrano, Marián Boguná, and Alessandro Vespignani. Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, 106(16):6483–6488, 2009.

**52.** Jari Saramäki, Mikko Kivelä, Jukka-Pekka Onnela, Kimmo Kaski, and Janos Kertesz. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, 75(2):027105, 2007.

**53.** Ernesto Estrada, Naomichi Hatano, and Michele Benzi. The physics of communicability in complex networks. *Physics Reports*, 514(3):89–119, 2012.