Characterising heavy-tailed networks using q-generalised entropy and q-adjacency kernels

Koponen, Ismo T.

2021-03-15

Characterising heavy-tailed networks using q-generalised entropy and q-adjacency kernels

Ismo T. Koponen *, Elina Palmgren, Esko Keski-Vakkuri

Department of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland

Abstract

Heavy-tailed networks, which have degree distributions characterised by slower than exponentially bounded tails, are common in many different situations. Some interesting cases, where heavy tails are characterised by inverse powers λ in the range $1 < \lambda < 2$, arise for associative knowledge networks, and semantic and linguistic networks. In these cases, the differences between the networks are often delicate, calling for robust methods to characterise the differences. Here, we introduce a method for comparing networks using a density matrix based on q-generalised adjacency matrix kernels. It is shown that comparison of networks can then be performed using the q-generalised Kullback-Leibler divergence. In addition, the q-generalised divergence can be interpreted as a q-generalised free energy, which enables the thermodynamic-like macroscopic description of the heavy-tailed networks. The viability of the q-generalised adjacency kernels and the thermodynamic-like description in characterisation of complex networks is demonstrated using a simulated set of networks, which are modular and heavy-tailed with a degree distribution of inverse power law in the range $1 < \lambda < 2$.

Key words: Heavy tailed networks, adjacency kernels, q-entropy, generalised thermodynamics

* Corresponding author
  Email address: ismo.koponen@helsinki.fi (Ismo T. Koponen).

22 December 2020

1 Introduction

When associative knowledge or term associations between words and terms are represented as networks, the networks are often found to be heavy-tailed [1–3]. Being heavy-tailed means here that the trailing edge of distribution of low frequency for node degrees d is not exponentially bounded but instead exhibits an inverse power-law type of decay $P(d) \propto d^{-\lambda}$ with $\lambda \in \,]1,3]$. In addition, such networks may also have a clear modular or community structure, originating, for example, from the thematic content of nodes [4]. For example, networks representing students' associative knowledge have distributions of degree and Katz centralities which are heavy-tailed and can be fitted with inverse power laws with exponents in the range of 1 to 2 [1,2]. The heavy tails with such values of λ also appear to be ubiquitous in word frequency, ranking distributions and many linguistic networks [3,5]. Powers from 1.5 to 2 are encountered in information and knowledge networks like Wikipedia and in semantic networks resulting from acquiring knowledge (terms) from large knowledge networks [5,6].

The research on associative networks often focuses on finding individual key items in the networks and their connectivities. Communicability centrality, as introduced by Estrada, and the closely related Katz centrality have been useful in exploring the connectivities of the nodes through long paths (or walks) [1,2]. In practice, the real networks in such applications are always quite small (about 1000-1500 nodes) and centrality measures have usually very large statistical variation. Therefore, it would be advantageous to seek for a descriptions which use all available information of the network states. However, taking the networks as holistic systems, that is, as connected sets of items, requires different approaches based on the description of the networks on macro-level [7,8]. For this, matrix transformations based on a complete adjacency matrix that describes the network are a promising starting point. For example, the communicability centrality is related to a matrix transformation which is obtained as an exponential transformation, while Katz centrality is related to Neumann transformation [9,10]. Here, we introduce a q-generalised [11–13] matrix transformation (called a q-adjacency kernel in what follows), which interpolates between these two well-known transformations. It is shown that the q-adjacency kernel provides a starting point for a thermodynamic-like macroscopic description of heavy-tailed networks and allows us to define a quantity that can be interpreted as the q-generalised free energy of a network.

The q-generalised free energy for a heavy-tailed network is obtained here by a derivation based on q-generalised information theoretic entropy. The information theoretic entropic measures have recently found many application in the characterisation of complex networks, multilayer and multiplex networks [7,8,14,15]. For example, the Jensen-Shannon divergence, which can be seen

2

as a symmetrised and regularised form of the Kullback-Leibler divergence, has proved to be a flexible and robust information theoretic entropic measure for comparison of the networks, because it is symmetric, bounded and it provides connection points to several related divergence measures within statistical physics [14,15]. The quantum version of the Jensen-Shannon divergence can be utilised effectively to measure the similarity between networks, as based on quantum walks [16]. Another class of information theoretic entropies is provided by Rényi and Tsallis entropies [8,11,17,18] which are closely related classes of entropies. These entropies belong to a class of q-generalised information theoretic measures, which have recently found applications in many complex systems, quantum systems and entangled systems [8]. The Rényi and Tsallis entropies form a family of information theoretic measures that generalise the Shannon entropy, and can be used as measures of mutual information and relative entropies (see e.g. [17]). Regarding the characterisation of heavy-tailed networks, the generalised q-entropies have the advantage of providing a parameter (q-index) to tune for specific properties of heavy-tailed networks, either for the low or high frequency part of the distributions, thus allowing foraging for information of the desired properties of the distribution [19–21]. Consequently, the Tsallis and Rényi entropies have found applications in linguistic and semantic networks [19] as well as in comparisons of texts based on word frequency occurrence [20,21], all these are cases where heavy-tailed networks are of interest. Here, we show that the Tsallis entropy and the q-generalised Kullback-Leibler (termed Kullback-Leibler-Tsallis in what follows) divergence [22,23] (see also [24,25]) are particularly suitable as a starting point to derive a q-generalised free energy to characterise the heavy-tailed network, thus providing access to a macro-level, thermodynamic-like description of the heavy-tailed networks.

To test the applicability of q-adjacency kernels for networks that are heavy-tailed and modular, such as associative knowledge networks, we use a recently introduced generative model to produce networks that resemble the empirical networks [26]. The minimal model to generate the networks is introduced first, in section 2. In section 3 we discuss in detail how q-generalised adjacency kernels are obtained, and how the thermodynamic-like interpretation is based on q-generalised Kullback-Leibler-Tsallis divergence. The section 3 ends with a discussion on how by using ensembles of the simulated heavy-tailed networks the validity of the derived theoretical results can be tested. Section 4 provides the results showing that the heavy-tailed networks yield to the thermodynamic-like macro-scale characterisation, based on the q-generalised entropy and free energy. In addition, Appendix A contains a sequel result of practical utility, for comparison of heavy-tailed networks. Finally, the implications and the practical use of the results are discussed.

3

2   Minimal generative model and simulations

Many real networks have hub-like nodes so that their degree distributions can be characterised as heavy-tailed distributions. Here, heavy tails refer to a form of the trailing edge of a distribution that decays considerably more slowly than a normal distribution, or to exponential distributions for which the trailing edge resembles to some degree the inverse power-law type distribution given as

$$P(d) \propto d^{-\lambda'}, \quad \lambda' \in \,]1,3]. \qquad (1)$$

The inverse power-law type distributions are used here as a model of heavy-tailed distributions. However, it is not assumed that the heavy-tailed distributions are genuinely inverse power-law or can be strictly characterised as such, since this rarely seems to be the case [27]. Nevertheless, for practical reasons, it is useful to consider the networks with heavy-tailed distributions using the model of inverse power-laws, because many essential characteristics of the networks are captured by that class of distributions [28]. In addition to heavy tails, many real networks are also highly modular, with the modularity emerging e.g. from content based features [4] or thematisation [2]. Here, we focus on empirically interesting cases of associative knowledge networks with $\lambda' \in \,]1,2]$ with high modularity Q (as defined following Newman [29]) within range $0.8 < Q < 0.9$ [1,2].

We utilise here the generative model to produce a set of heavy-tailed networks that have a highly modular structure [26]. The model is based on generation of affinities for a fixed set of nodes, with the nodes assigned to different, pre-fixed classes. Here, we use parameters which generate networks of approximately N=1000 connected nodes and with inverse powers λ of degree-distributions obtaining values λ =1.3, 1.5, 1.7 and 1.9. The networks are generated using, for affinities π_k of nodes, the distribution

$$P(\pi_k) = P_0 \left[1 - (\lambda - 1)\Lambda \pi_k\right]^{-1/(\lambda-1)}, \quad \lambda \in \,]1,2], \quad \Lambda > 0. \qquad (2)$$

It is interesting to note that Eq. (2) can also be written in the form of a q-exponential $P(\pi_k)/P_0 = \exp_q[\Lambda \pi_k]$, which is the q-generalisation of the exponential function [11–13]. The parameter λ determines the inverse power of the resulting degree distribution of the network, while Λ controls the cut-off of affinities, and $\pi_k < \Lambda(\lambda-1)$. The affinity distribution can be derived in several ways and by using several parametrisations, [26,30,31]. What is essential here is that the affinity distribution in Eq. (2) can be used to generate networks with the desired inverse power. Due to modularity, however, the resulting inverse power of degree distribution is never exactly λ but usually somewhat larger, $\lambda' > \lambda$ [26].

In simulations, only one nested modular structure is used, which is three-tiered; the smallest modules contain N' potential nodes, the second tier three

4

of these smaller modules and thus 3xN' , and the highest and third tier contains three of the modules of 3xN' nodes, thus forming a set of N=3x(3xN') nodes in total. The connections between N nodes are established probabilistically as governed by the affinity distribution. Therefore, not all nodes in the nested tiers will be connected. With parameters chosen so that the resulting network has the size of about N=1000 connected nodes, the inverse power in range $1.3 < \lambda < 2.0$ and the modularity in range $0.8 < Q < 0.9$. The modularity of the networks is relaxed step-by-step, by rewiring until no changes due to rewiring are observed. In the rewiring, the degree sequence of the nodes is preserved. We have shown that such affinity-based linking of nodes with minimal other assumptions reproduces the empirically discovered properties of associative knowledge networks quite adequately [26].

The simulations to generate networks and their analysis are carried out in `Mathematica` using the `IGraph` package [32]. `IGraph` provides functionality for generating efficiently affinity-based networks simply by providing the probabilities π_k for the routine `IGStaticFittnessGame`. The output of the routine is a network with a predetermined number of links, linked according to the probabilities π_k drawn from distribution $P(\pi_k)$ in Eq. (2). The relaxation of modularity by rewiring is done using routine `IGRewire`. For each network, the simulations provide the adjacency matrix $\mathbf{A}$, which is then used to obtain the desired adjacency kernels to characterise the networks. Next, we turn to how the adjacency kernels are constructed and how the heavy-tailed networks can be characterised using them.

3   Thermodynamic-like description of networks with heavy tails

The thermodynamic-like description of the networks with heavy tails can be constructed using a suitable generalisation of the matrix transforms that have been used more traditionally to describe and characterise networks. It is shown how these generalisations provide a basis for constructing the thermodynamics-like description and and generalised free energy -like quantity for characterisation of the network. In that, the discussion closely follows the study by Abe and Rajagopal [33], who derived the q-generalised thermodynamics for q-generalised ensembles. The motivation of [33] was to develop a non-extensive version of thermodynamic laws for finite systems where the infinite volume thermodynamic limit is not applicable and surface effects violate extensivity. Similar constraints apply to finite complex networks, hence we are interested in exploring connections to non-extensive thermodynamics by applying the results and approaches of [33] to networks with heavy tails.

On the most basic level, networks are described by the set of nodes (vertices) and links (edges) between the nodes. This information is provided by the network's adjacency matrix $\mathbf{A}$, whose element $a_{ij} = [\mathbf{A}]_{ij}$ has the value 1 if the nodes i and j are connected. Otherwise its value is 0. From the adjacency matrix, many other relevant properties of the network become available through matrix transformations. Following previous studies on adjacency kernels [9,10] we define a graph kernel as a symmetric and positive semidefinite function which denotes a certain similarity or proximity between two nodes in the graph. All such graph kernels of interest here are based on the adjacency matrix $\mathbf{A}$ of the graph and are thus called adjacency kernels. Two well-known adjacency kernels, the exponential and the Neumann kernel, are related to path counting between nodes, thus providing information on how the nodes are globally connected [9,10].

The exponential kernel is one of the matrix transformations most widely used to analyse networks. The exponential kernel is defined as [9,10,34–38]

$$\exp[\beta\mathbf{A}] = \mathbf{I} + \frac{\beta^1 \mathbf{A}^1}{1!} + \frac{\beta^2 \mathbf{A}^2}{2!} + \frac{\beta^3 \mathbf{A}^3}{3!} + \dots, \qquad (3)$$

where matrix power $\mathbf{A}^k$ counts paths (walks) of length k. The exponential adjacency kernel weights the paths according to their length k by a weight factor β^k and inverse of the factorial $k!$. Therefore, connections with short paths gain more importance than those with long paths. The diagonal elements of the exponential kernel are called the Estrada index and the so-called communicability centrality is obtained as a row sum of the off-diagonal elements [35]. The exponential kernel and the associated measures (the Estrada index and the communicability centrality) perform well in finding globally important nodes. In addition, the exponential kernel is computationally very robust and stable, which is an advantage for analysis of complex networks [34–36,39].

The Neumann kernel [9,10] is another well-known adjacency kernel with many applications in characterising complex networks [40,41]. The Neumann kernel is given by [9,10]

$$[\mathbf{I} - \beta\mathbf{A}]^{-1} = \mathbf{I} + \beta^1 \mathbf{A}^1 + \beta^2 \mathbf{A}^2 + \beta^3 \mathbf{A}^3 + \dots, \qquad (4)$$

where the weight factor β must be chosen so that β^{-1} is larger than the largest eigenvalue of matrix $\mathbf{A}$. The row sum of the off-diagonal elements of the ith row in the Neumann kernel provides the Katz centrality [40,41] of node i, widely applied in social network analysis in finding the influential nodes [42]. The Katz centrality has also been successfully used in finding the central key nodes in heavy-tailed associative networks, which are of interest here [2]. One

disadvantage of the Katz centrality is that it is difficult to optimise the weight factor β to maintain the stability of computation and the desired resolving power [10,43]. On basis of effects of the choice of parameters in behaviour of the kernels and consequently, on ranking of nodes, it seems to be advisable to strive for largest values of weight factor but to check how results depend on the choice of it [10].

The generalisation of the walk counting, which interpolates between the exponential and Neumann kernels, is the q-generalised exponential kernel [33]

$$\exp_q[\beta\mathbf{A}] = [\mathbf{I} - (q-1)\beta\mathbf{A}]^{-1/(q-1)}, \quad q \in \,]1,2] \text{ and } \beta > 0, \qquad (5)$$

where β must be chosen so that $\beta < 1/B$ with B being the largest eigenvalue of matrix $\mathbf{A}$. The q-exponential adjacency kernel can be expanded in series

$$\exp_q[\beta\mathbf{A}] = \mathbf{I} + \beta^1\mathbf{A}^1 + \frac{q}{2!}\beta^2\mathbf{A}^2 + \frac{2q^2-q}{3!}\beta^3\mathbf{A}^3 + \ldots = \sum_k c_k(q)\,\beta^k\mathbf{A}^k \qquad (6)$$

$$c_k(q) = (1-q)^k \binom{\frac{1}{1-q}}{k} = \frac{(1-q)^k}{k!}\left(\frac{1}{1-q} + 1 - k\right)_k \qquad (7)$$

where $\binom{\cdot}{\cdot}$ is the binomial coefficient and $(\cdot)_k$ is the Pochhammer ascending factorial [44]. In the limit $q \to 1$ the q-adjacency kernel approaches the exponential kernel since $c_k \to 1/k!$ while for $q \to 2$ it approaches the Neumann kernel when $c_k \to 1$. The q-adjacency kernel interpolates between walk counting where walks of length k are either weighted by β^k (Neumann kernel) only or much more heavily with $\beta^k/k!$ (exponential kernel). The q-adjacency kernel is also a positive semidefinite matrix and has non-negative eigenvalues when β is smaller than the inverse of the largest eigenvalue of the adjacency matrix. The q-adjacency kernel in Eq. (5) opens up a path to two interesting generalisations. First, the q-kernel provides a starting point for constructing the thermodynamic-like description of the networks with a heavy-tailed degree distribution. Second, it leads to q-generalisations of the Estrada-type communicability centrality, which is of high practical utility in characterising heavy-tailed networks. However, while it is useful, we have not found a thermodynamic-like interpretation for it. Hence we have relegated the discussion of the generalised communacibility centrality to Appendix A. We proceed next to the details of the thermodynamic-like interpretation of the kernels.

3.2   Thermodynamic-like description of heavy-tailed networks

The heavy-tailed networks generated by simulations and based on the generic model introduced in section 2 are described using adjacency matrix $\mathbf{A}$. Using

the q-adjacency kernels we characterise the connectivity (or communicability) properties of the network by defining the density matrix

$$\boldsymbol{\rho} = Z_q^{-1}[\mathbf{I} - (q-1)\beta\mathbf{A}]^{-1/(q-1)} \equiv Z_q^{-1}\exp_q(\beta\mathbf{A}), \tag{8}$$

where $Z_q = \mathrm{Tr}[\mathbf{I} - (q-1)\beta\mathbf{A}]^{-1/(q-1)}$ is the q-generalised partition function (compare with [33]). The density matrix provides a complete description of the network and, in that sense, is a holistic measure for the characterisation of networks. Moreover, it can be taken as a starting point for robust comparison of the structure and structural similarity of networks [7,8].

3.2.1 Thermodynamics corresponding to Gibbs states with $q = 1$

We review first a thermodynamic interpretation of the density matrix in the case $q = 1$, where

$$\boldsymbol{\rho} = Z^{-1}\exp(\beta\mathbf{A}) . \tag{9}$$

There are two obvious ways to interpret (9) as a thermal (Gibbs) state. First, in a system with a Hamiltonian $\mathbf{H} = -\mathbf{A}$,

$$\boldsymbol{\rho} = Z^{-1}\exp(-\beta\mathbf{H}) . \tag{10}$$

with a positive inverse temperature $\beta > 0$, where the prefactor $Z = Z(\beta) = \mathrm{Tr}(e^{-\beta\mathbf{H}})$ is the partition function. For example, one could consider the Hamiltonian as the graph Laplacian $\mathbf{H} = \mathbf{D} - \mathbf{A}$ without the diagonal matrix $\mathbf{D}$ of the node degrees (compare with [7,8]). Second, one can choose the adjacency matrix directly as the Hamiltonian, $\mathbf{H}' = \mathbf{A}$ (the prime is used to distinguish this alternative choice), but then one must choose a negative temperature $\beta' = -\beta < 0$ to write $\exp(\beta\mathbf{A}) = \exp(-\beta'\mathbf{H}')$ [34,36]. Negative temperatures arise naturally in systems with a bounded spectrum of energy eigenvalues (as in this case, $\mathbf{H}$ has a finite set of eigenvalues). For convenience, we proceed with the first interpretation with the Hamiltonian $\mathbf{H}$.

The Gibbs state maximizes the von Neumann entropy

$$S(\boldsymbol{\sigma}) = -\mathrm{Tr}(\boldsymbol{\sigma}\log\boldsymbol{\sigma}) \tag{11}$$

with the constraint that the expectation value of the Hamiltonian $\langle\mathbf{H}\rangle_{\boldsymbol{\sigma}} = \mathrm{Tr}(\boldsymbol{\sigma}\mathbf{H})$ is kept fixed. That is, with the constraint, the density matrix $\boldsymbol{\sigma}$ for which $S(\boldsymbol{\sigma})$ is maximised, is the Gibbs state $\boldsymbol{\sigma} = \boldsymbol{\rho}$. Substituting the Gibbs state to the von Neumann entropy, so that it becomes the thermal entropy $S(\beta) = S(\boldsymbol{\rho})$ of the ensemble, the first law of thermodynamics holds,

$$F(\beta) = E(\beta) - \beta^{-1}S(\beta) , \tag{12}$$

with the free energy $F(\beta) = -\beta^{-1}\log Z(\beta)$ and thermal energy $E(\beta) = \langle H\rangle_{\boldsymbol{\rho}}$.

8

We now move to consider comparisons between density matrices $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$. Suppose $\boldsymbol{\sigma} = Z^{-1}(\beta)\exp(-\beta\mathbf{H})$ and $\boldsymbol{\rho} = \boldsymbol{\sigma} + \delta\boldsymbol{\rho}$ by a small perturbation $\delta\boldsymbol{\rho}$ with $\mathrm{Tr}(\delta\boldsymbol{\rho}) = 0$. We define the change in entropy

$$\delta S \equiv S(\boldsymbol{\rho}) - S(\boldsymbol{\sigma}) = S(\boldsymbol{\rho}) - S(\beta) \tag{13}$$

and a change in energy

$$\delta E \equiv \langle\mathbf{H}\rangle_{\boldsymbol{\rho}} - \langle\mathbf{H}\rangle_{\boldsymbol{\sigma}} = \mathrm{Tr}(\boldsymbol{\rho}\mathbf{H}) - E(\beta) \ . \tag{14}$$

Then, it is known that the relative entropy (Kullback-Leibler divergence) between $\boldsymbol{\rho}, \boldsymbol{\sigma}$,

$$K[\boldsymbol{\rho}||\boldsymbol{\sigma}] = \mathrm{Tr}[\boldsymbol{\rho}(\ln\boldsymbol{\rho} - \ln\boldsymbol{\sigma})] \tag{15}$$

for the infinitesimal variation $\boldsymbol{\rho} = \boldsymbol{\sigma} + \delta\boldsymbol{\rho}$ from the thermal density matrix $\boldsymbol{\sigma}$ gives an infinitesimal change in free energy δF with (compare with [34,36])

$$\beta\delta F = K[\boldsymbol{\rho}||\boldsymbol{\sigma}] = \delta E - \beta^{-1}\delta S \ . \tag{16}$$

Next we proceed to show that by analogous reasoning similar results can be obtained for q-generalised states.

3.2.2   *Thermodynamics corresponding to q-generalised states with $q > 1$*

It is possible to generalise the above derivation based on Gibbs states with $q = 1$ for q-generalised states corresponding to the index $q > 1$. One starts with the constrained extremisation problem, now with the goal of maximising the generalised q-entropy (Tsallis entropy) of form [11,22,23,33,45,46]

$$S_q[\boldsymbol{\sigma}] = \frac{1}{1-q}(\mathrm{Tr}\boldsymbol{\sigma}^q - 1). \tag{17}$$

In the limit $q \to 1$ the q-entropy reduces to the usual von Neumann entropy (11). Now we add a constraint that the expecation value $\langle\mathbf{H}\rangle_{\boldsymbol{\sigma}} = \mathrm{Tr}(\boldsymbol{\sigma}\mathbf{H})$ is kept fixed and ask what is the form of the density matrix $\boldsymbol{\sigma}$ that maximises entropy in Eq. (17). The result turns out to be $\boldsymbol{\sigma} = \boldsymbol{\rho}_q$, with [33,45]

$$\boldsymbol{\rho}_q = Z_q^{-1}\exp_q(-\beta\mathbf{H}) \ , \tag{18}$$

where $Z_q = Z_q(\beta) = \mathrm{Tr}[\exp_q(-\beta\mathbf{H})]$. Note that setting $\mathbf{H} = -\mathbf{A}$, Eq. (18) agrees with Eq. (8) as expected. This gives a q-generalised density matrix in Eq. (8) for description of the networks with adjacency matrix $\mathbf{A}$.

Next we will consider two closely related networks, the original modular one with a state $\boldsymbol{\sigma}$ and the rewired counterpart with $\boldsymbol{\rho} = \boldsymbol{\sigma} + \delta\boldsymbol{\rho}$. The original state is q-distributed, with $\boldsymbol{\sigma}$ of the form in Eq. (8). We are interested in finding the q-generalised relation of (16) between changes in free energy, energy,

and entropy. The q-distribution arises from constrained maximisation of the Tsallis q-entropy, generalising the von Neumann entropy. Correspondingly, the appropriate quantity that generalises the Kullback-Leibler divergence is the q-generalised Kullback-Leibler-Tsallis divergence (henceforth q-divergence for brevity), given by [22,23,33,45]

$$K_q[\boldsymbol{\rho}||\boldsymbol{\sigma}] = \frac{1}{1-q}[1 - \mathrm{Tr}(\boldsymbol{\rho}^q \boldsymbol{\sigma}^{1-q})], \tag{19}$$

which at the limit $q \to 1$ approaches the Kullback-Leibler divergence. The q-divergence is always positive, and zero for identical states. Note that when $q \in]1, 2]$ and density matrices are positive semidefinite, the term $\boldsymbol{\sigma}^{1-q}$ needs to be interpreted as $[\boldsymbol{\sigma}^{-1}]^{q-1}$, where $\boldsymbol{\sigma}^{-1}$ is the pseudo-inverse of density matrix $\boldsymbol{\sigma}$. Here, we have followed the ordering of density matrices $\boldsymbol{\rho} \leq \boldsymbol{\sigma}$ so that states of $\boldsymbol{\rho}$ (corresponding to the rewired networks) are obtained from $\boldsymbol{\sigma}$ (corresponding to the original modular network), meaning that the number of links in the rewired network is always equal or smaller than in the original one.

For the q-generalisation of the infinitesimal first law in Eq. (16), we follow [33]. We first need a q-generalisation of the expectation value of the Hamiltonian, defined by [33]

$$\bar{E}_q = \mathrm{Tr}(\mathbf{H}\boldsymbol{\rho}^q)/\mathrm{Tr}(\boldsymbol{\rho}^q) , \tag{20}$$

the trace in the denominator is needed for an expectation value with respect to a unit normalised density matrix $\boldsymbol{\rho}^q/\mathrm{Tr}(\boldsymbol{\rho}^q)$. With this definition, by a direct calculation one obtains [33] an extension of the usual thermodynamic relation,

$$\delta S_q/\delta \bar{E}_q = \beta, \quad q \in]1, 2] \text{ and } \beta > 0 , \tag{21}$$

in agreement with the interpretation of β as the inverse temperature. Starting from the q-divergence as defined in Eq. (19) and assuming that changes $\delta\boldsymbol{\rho}$ in state $\boldsymbol{\rho}$ are small enough so that $\mathrm{Tr}\delta\boldsymbol{\rho} \approx 0$ and $\boldsymbol{\rho} \approx \boldsymbol{\sigma}$ in all cases, we can show by direct calculation (compare with derivation in ref. [33]) that

$$(\mathrm{Tr}\boldsymbol{\sigma}^q)\,\delta K_q[\boldsymbol{\rho}||\boldsymbol{\sigma}] = \beta\,\delta \bar{E}_q - \delta S_q[\boldsymbol{\rho}], \quad q \in]1, 2]. \tag{22}$$

It is now possible to interpret the term $(\mathrm{Tr}\boldsymbol{\sigma}^q)\,\delta K_q[\boldsymbol{\rho}||\boldsymbol{\sigma}]$ as analogous to a change in free energy of a thermodynamic systems, by defining a change in q-free energy as

$$\delta F_q = \beta^{-1}(\mathrm{Tr}\boldsymbol{\sigma}^q)\,\delta K_q[\boldsymbol{\rho}||\boldsymbol{\sigma}]. \tag{23}$$

Then, a relation resembling the infinitesimal form of the first law of thermodynamics holds,

$$\delta F_q = \delta \bar{E}_q - \beta^{-1}\delta S_q. \tag{24}$$

We now restore the adjacency matrix by replacing $\mathbf{H} = -\mathbf{A}$, and denote its q-expectation value by $\bar{A}_q = \mathrm{Tr}(\mathbf{A}\boldsymbol{\rho}^q)/\mathrm{Tr}(\boldsymbol{\rho}^q)$, so that $\delta \bar{E}_q = -\delta \bar{A}_q$. The

generalised first law then reads

$$\delta F_q = -\delta \bar{A}_q - \beta^{-1} \delta S_q \ . \tag{25}$$

In the alternative interpretation, where the adjacency matrix is chosen as the Hamiltonian, $\mathbf{H'} = \mathbf{A}$, we would have arrived at this form. The minus sign in front of $\bar{A}_q$ is consistent with the inverse temperature $\beta' = -\beta$ being negative in the alternative interpretation. In that case, the alternative form of (24) is rewritten as

$$\delta F'_q = \delta \bar{E}'_q - (\beta')^{-1}\delta S_q, \tag{26}$$

where we have defined $\delta F'_q = (\beta')^{-1}(\mathrm{Tr}\boldsymbol{\sigma}^q)\, \delta K_q[\boldsymbol{\rho}||\boldsymbol{\sigma}] = -\delta F_q$ as a change in q- free energy.

At the limit $q \to 1$ the result in Eq. (25) agrees with the results and interpretation suggested by Estrada [34–36], who has proposed a similar relation by using the exponential adjacency kernel and Kullback-Leibler divergence. The result in Eqs. (22)-(25) can be taken as q-generalisation of the previous results by Estrada. The advantage of the q-generalised adjacency kernels is that using the freedom provided by choice of parameters β and q, the contribution of paths of different lengths to the connectivity (or communicability) of nodes can be explored. At the limit $q \to 1$ short paths are more dominant for connectivity than long ones, while at the limit $q \to 2$ the contribution of longer paths becomes more important. In both cases, also weight factors can be used to tune the scale of paths that one wishes to take into account in the connectivity of nodes. At the limit $\beta \to 0$, a vanishing weight is attached to all links, so that the system is essentially a set of disconnected nodes, while the larger the value of β, more tightly connected the system.

3.3 Generalised free energy and response function from simulations

The viability of the thermodynamic-like description of heavy-tailed networks is tested by first generating a set of networks using the generative model, designed to produce networks with inverse power law type degree distribution with power $1 < \lambda < 2$ and with high modularity. To obtain the q-generalised free energy from simulations is, however, not entirely straightforward for two reasons. First, the systems are always finite and thus the density matrix as defined in Eq. (8) does not exactly maximise the q-entropy. Secondly, due to the small size of the systems, statistical fluctuations are large, and it is difficult to obtain the quantities $\mathrm{Tr}\boldsymbol{\sigma}^q$, δS_q and $\delta \bar{A}_q$ accurately, while change in divergence δK_q is computationally more stable. Therefore, based on the simulations where networks are generated, we first justify that relation in Eq. (22) holds by demonstrating the validity of equivalence

$$K_0(\beta)\, \delta \tilde{K}_q = -\beta A_o(\beta)\delta \bar{A}_q - \delta S_q. \tag{27}$$

11

Here we have assumed that $K_0 \simeq \mathrm{Tr}\boldsymbol{\sigma}^q \mathrm{Max}[\delta K_q]$ depends only on β, and $\delta \tilde{K}_q = \tilde{K}_q / \mathrm{Max}[\tilde{K}_q]$ only on rewiring frequency π. The normalisation parameter $A_0(\beta)$ is allowed to compensate for the β-dependence of the normalisation factor $\mathrm{Tr}\boldsymbol{\rho}^q$, due to the finite size of networks affecting the cut-off of eigenvalues of the adjacency matrix. The validity of Eq. (22) is then guaranteed in the region of parameters where the scaling functions K_0 and $\tilde{K}_q$ can be found so that Eq. (27) holds. As will be seen, this is possible for networks with heavy tails characterised by $1 < \lambda < 2$ when $1.3 < q < 1.7$ (in some cases also for lower and higher values of q). In this region of parameter q, the thermodynamic-like description is viable and the q-generalised free energy change is

$$\delta F_q = \beta^{-1} K_0 \, \tilde{K}_q, \tag{28}$$

where functions K_0 and $\tilde{K}_q$ are obtained by fitting to simulation data.

The q-generalised free energy allows us to define how the network responds to changes in the external parameters β and π. In the thermodynamic interpretation, the response to changes in β is analogous to heat capacity

$$\chi = \beta^2 \frac{\partial^2 (\beta F_q)}{\partial \beta^2} \propto \beta^2 \frac{\partial^2 K_0}{\partial \beta^2}, \tag{29}$$

where it is assumed that the scaling with regard to β is the same for the rewired and non-rewired (fully modular) networks. Then, the response is determined up to an unknown constant (only the change in free energy due to rewiring is known). The response function allows a simple description how and in what state the network is most responsive to changes in the strength of the links. Different definitions of the response function are possible (see e.g. [47]), but here we have preferred to retain the analogy with a thermal response (i.e. heat capacity).

4 Results and discussion

The results for the diagonal values of the q-adjacency kernels are discussed first, because they have the closest relation to a thermodynamic-like interpretation in the case $q = 1$, when diagonal values provide the so-called Estrada index, which yields to a thermodynamic interpretation in terms of free energy [34,36]. In small systems, however, the distribution of eigenvalues has so large fluctuations that the utility of such interpretation is in practice diminished. Here, we show results that demonstrate that a thermodynamic-like interpretation similar to that proposed by Estrada for diagonal components is viable also for generalised q-adjacency kernels, which describe the state of the network as a whole, but with much reduced fluctuations.

The simulation results to be discussed here show that within a certain region of parameters q and β we can maintain the interpretation of divergence difference ΔK_q as change of free energy ΔF_q of the network (note that for simulation results, symbols $\Delta(\cdot)$ instead of $\delta(\cdot)$ without subscript q are used). One of the central choices is the choice of β_{MAX}, which must be lower than the inverse of the highest eigenvalue of the adjacency matrix. Since the highest eigenvalue of the adjacency matrix depends on the details of the network, the exact choice of maximum value is case-dependent. In practice, however, for the heavy-tailed networks of interest here, values of β which are from 50% to 70% of the maximum value β_{MAX} provide the best compromise between stability of results and resolving power [2]. On the other hand, when β is too low, resolving power with regard to differences in characteristics of nodes is lost. Therefore, the β_{MAX} is first chosen to be about 70% of the highest possible value, and it depends on λ linearly so that $\beta_{\mathrm{MAX}} = -0.02 + 0.33\lambda$ for $\lambda \in]1, 2]$. The values of β_{MAX} for different choices of λ are summarised in Table I. The adjacency kernel for each value of λ is then evaluated for seven values ranging from β_{MAX} to $0.7\beta_{\mathrm{MAX}}$.

Table I
The parameters β_{MAX} in simulations and the maximum number of attempted links M_{A} and realised links M_{R} (on average) with M_0=2000. In all cases, the number of connected nodes is on average about N=1000. The realised powers λ' for given λ are also shown.

	$\lambda = 1.3$	$\lambda = 1.5$	$\lambda = 1.7$	$\lambda = 1.9$
β_{MAX}	0.023	0.030	0.036	0.043
M_{A}/M_0	2.0	1.4	1.1	1.0
M_{R}/M_0	1.8	1.3	1.0	0.9
λ'	1.7	1.8	1.9	2.1

In the simulations to produce generic networks, only one type of modularity is chosen, corresponding to the three-tiered modular structure where the first tier has units of N'=200 nodes, the second tier three of such units of 200 nodes (for a total of 600), and the third tier three units of 600 nodes, providing N=1800 nodes in total. All nodes, however, are not connected and the choice of parameters lead to total number of connected nodes of about N=1000. The structure is thus a block model with a three-tiered hierarchy of connections. This selection results in the initial modular structure with modularity in range $0.80 < Q < 0.90$ depending on the case. As was explained in section 2, the parameters are selected so that they roughly correspond to

values found empirically for networks of associative, thematic knowledge [1,2]. However, provided that the modularity is high enough, the exact structure of initial modularity and its variation are not crucial for the properties of interest here (for effects of modularity, see ref. [26]). Therefore one initial state is enough. Instead, the relaxation of the initial modularity and its effects are of interest. The modularity of the networks is relaxed by rewiring the links with relative frequency $\nu \in [0, \nu_{\mathrm{MAX}}/M]$. In practise, the value $\nu_{\mathrm{MAX}} = 5.0 \cdot 10^5$ is chosen to guarantee complete rewiring, while $\nu = 1$ means that, on average, each link is rewired once. In the rewiring, performed with `IGraph` algorithm `IGRewire`, the degree sequence of the nodes is preserved.

4.2 The diagonal values of density matrices

The degree centrality distributions $P(d)$ for degrees of nodes are shown in Figure 1 (the upper row). The distributions are clearly heavy-tailed and can follow a power law in a broad range of degrees. It should be noted that irrespective of the choice of λ, defining the affinity distribution, the powers corresponding to the degree distribution are rather narrowly distributed from $\lambda' \approx 1.7$ for $\lambda = 1.3$ up to $\lambda' \approx 2.0$ for $\lambda = 1.9$. This behaviour is most likely caused by the fact that due to constraints set by modularity, high affinity nodes are more likely to be connected when only about 1000 of all the potential 1800 nodes are connected. A similar effect of values λ' exceeding the values of λ was also observed in simulations where a broader set of modularities was explored [26].

The density distribution of diagonal values $[\boldsymbol{\rho}]_{ii}$ of density matrix $\boldsymbol{\rho}$ are related to the eigenvalues of the density matrix and are shown in two lower rows in Figure 1 for the fully modular and fully rewired networks. The diagonal values have different distributions corresponding to different values of q and thus show that they are more sensitive to the details of the networks than the degree distribution. Moreover, the relaxation of modularity substantially affects the leading edge of the distribution. The modular networks have high number of low values of $[\boldsymbol{\rho}]_{ii}$ due to modular structure, in which local connections in modules are tight. However, with relaxation of the modularity, the distributions become more power-law type and higher values of $[\boldsymbol{\rho}]_{ii}$ become more dominant, indicating that the number of longer paths increases with loss of modularity. This effect is more pronounced the higher the value of q. For the lowest values of q the distribution of the diagonal values of $[\boldsymbol{\rho}]_{ii}$ is in all cases close to an inverse power law distribution. For higher values of q, the high frequency leading edge corresponding to the low values of $[\boldsymbol{\rho}]_{ii}$ begins to deviate from power law behaviour. Also, in the low frequency part, the statistical fluctuations are high. In practice, for example in case of associative knowledge networks, just this part of the distribution is of interest. With small sample sizes, however, the large statistical fluctuations become an obstacle for
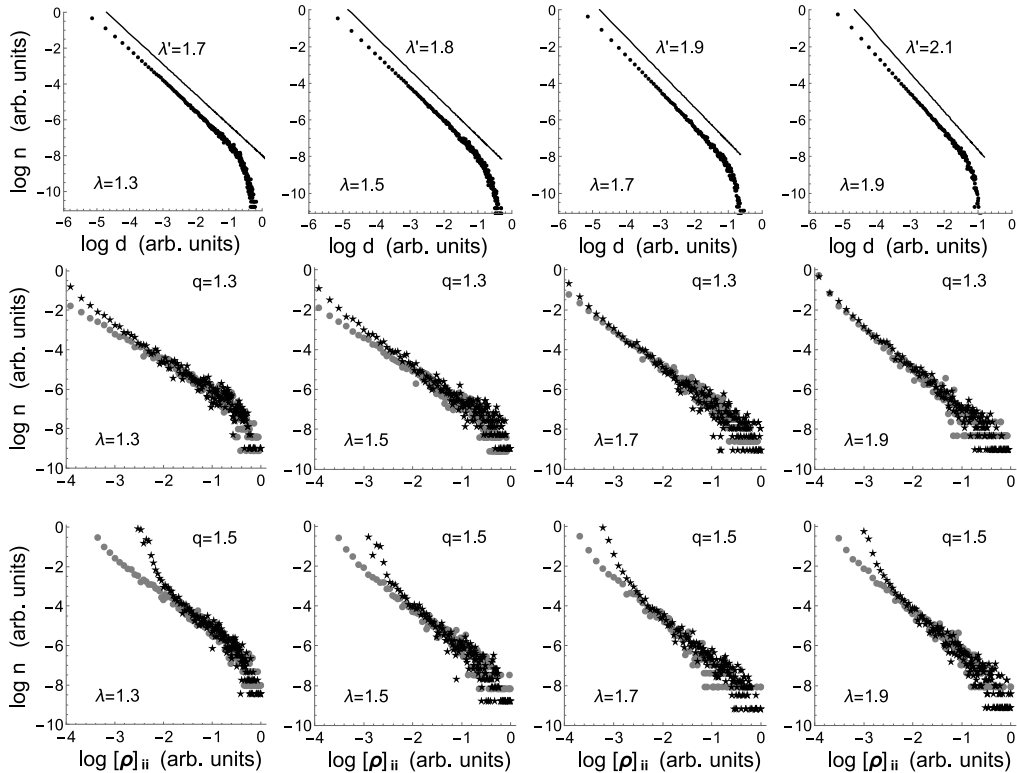
14

Fig. 1. Distributions of degree d for parameters $\lambda = 1.3$, 1.5, 1.7 and 1.9 (upper row, with estimates of power-law fits shown by λ'). The distributions of diagonal elements $[\boldsymbol{\rho}]_{ii}$ of density matrix $\boldsymbol{\rho}$ for each λ and values $q = 1.3$, 1.5 and 1.7 are shown on two lowest rows, respectively. In each case, the distributions are shown for original modular (symbol $\star$, upper sets of data points) and completely rewired networks (symbol $\bullet$, lower sets of data points) with $\beta = \beta_{\max}$. The results for degree distributions are with 10^4 repetitions while for $[\boldsymbol{\rho}]_{ii}$ 100 repetitions are used.

similarity comparisons based on Kullback-Leibler divergence; the statistical fluctuations are usually too large to allow robust results [2,26].

4.3 The thermodynamic-like interpretation

To compare heavy-tailed networks and the effects of modularity we use the q-generalised entropy, q-generalised divergences, and the complete density matrix $\boldsymbol{\rho}$ to characterise the networks. In addition, a suitably chosen q-adjacency kernel may allow macroscopic description based on q-generalised free energy. To secure this, we need to show that the relation in Eq. (27) holds for simulation results. Figure 2 shows the change ΔK_q (simulation results for δK_q and with subscript q omitted in what follows) for different values of q and β as a function of rewiring frequency ν so that all divergences are scaled to

15

the maximal value $K_0 = \text{Max}[\Delta K]$. The reference (density matrix $\boldsymbol{\sigma}$ in Eq. (19)) is chosen to be the original, fully modular network. The curves for ΔK are sigmoidal, but their maximum values $K_0 = \text{Max}[\Delta K]$ are different for different choices of q and β as shown in Fig. 2 in the middle panel. However, the scaled values $\Delta \hat{K} = \Delta K / K_0$ collapse to a single curve (shown at left in Fig. 2) that is the same for all choices of $q \in [1.3, 1.9]$ and $\beta/\beta_{\text{MAX}} \in [0.7, 1.0]$, thus demonstrating the scaling of the ΔK. The transitional region where the properties of the network change essentially is now recognised to be located in $-1 < \text{Log}\, \nu < 1$ and in region $\text{Log}\, \nu > 2$ (note that here Log refers to natural logarithm) the network is essentially completely rewired (i.e. random, only degree sequence preserved). The most interesting properties of the changes of the network, however, are contained to the maximal value $K_0(\beta; q, \lambda) = \text{Max}[\Delta K]$, which depends on β and parameters q and λ. This quantity is discussed in more detail below.

To demonstrate the validity of Eq. (27) the scaled values of simulation results for $\Delta K + \Delta S$ and $-\beta \Delta A$ (corresponding to quantities at left and right side in Eq. (27), respectively) are shown separately on the panel at right of Fig. 2. If Eq. (27) holds, the scaled curves should be linearly proportional. This is suggested by the appearance of the curves shown. The curves, however, are averaged for each set of parameters over the 100 repetitions where the statistical fluctuations are still large, as shown in Fig. 2. The Pearson correlation of data points for $\Delta K + \Delta S$ and $-\beta \Delta A$ is in all cases from 0.97 to 0.99 with p-values well below 10^{-4}, indicating good correlation in most cases, and at worse close to 10^{-3} for high values of q, still indicating a reasonable correlation. This means that we can take the relation in Eq. (27) to be valid in the cases studied. The dependence is also linear with good accuracy; the residuals of the averaged values from linear fits remain below 10% of the mean values for the average values shown in Fig. 2 for $q < 1.5$, increasing to about 30% in the worst cases for $q = 1.9, \lambda = 1.3$ and $\beta \approx \beta_{\text{MAX}}$. The results indicate that the interpretation of the divergence in terms of free energy as suggested by Eqs. (25) and (28) is viable provided that parameter q is in range $1.3 < q < 1.7$ and close enough λ for $q < 1.7$. Values $q < 1.3$ could not be tested, because computations rapidly become unstable for low values of q.

The dependence of the maximal value $K_0 = \text{Max}[\Delta K]$ on β and parameters q and λ is somewhat complex but regular. In Figure 3, at the left panel, K_0 is shown scaled in a suitable way, to obtain a linear dependence. The linear form $\kappa_0(\beta/\beta_{\text{MAX}}) = a + b(1 - \beta/\beta_{\text{MAX}})$ with $a = -1.86$ and $b = 47.1$ shown in Fig. 3 is obtained by scaling

$$\kappa_0 = \bar{\kappa}_\lambda K_0^{-\alpha_\lambda} - C_\lambda. \tag{30}$$

Here coefficients $\bar{\kappa}_\lambda(q)$ and $C_\lambda(q)$ and power $\alpha_\lambda(q)$ depend on parameters λ and q as shown in Fig. 3 in the middle panel. The sigmoidal fitting functions
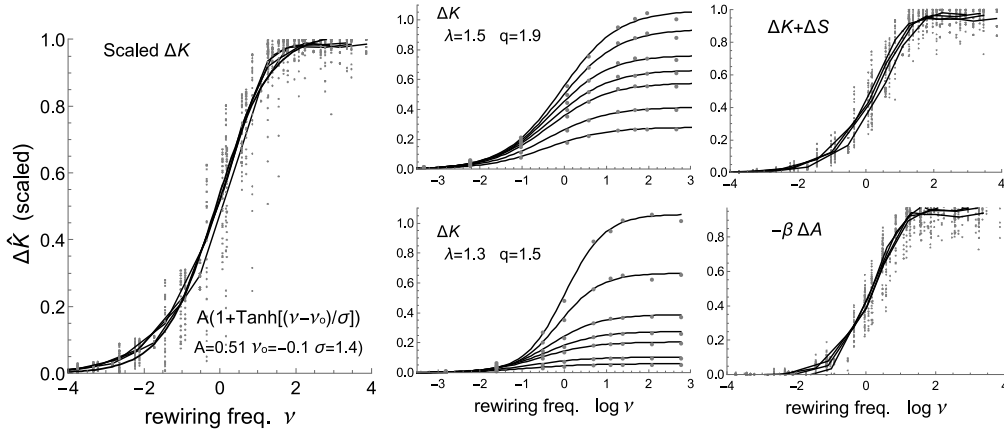
Fig. 2. The scaled divergence difference $\Delta\hat{K}$ (see text for definition) is shown on the left panel as a function of rewiring frequency ν. In the middle panel are shown values of ΔK (scaled to maximal value in the ensemble) for heavy-tailed networks with different values of λ. At right are shown the (scaled) values of $\Delta K + \Delta S$ (the upper figure) and $-\beta\Delta A$ (the lower figure). Note that interpretation of ΔK as free energy requires that these terms are linearly proportional. In all cases, data-points are plotted showing the variance in data. The natural logarithm Log is used throughout.

displayed in the figures for these parameters are obtained from simulation results through simple, descriptive fits to data. It should be noted that there is no theoretical basis to motivate the form of the fits so they are only practical vehicles for interpolation and moderate extrapolations. Moreover, the number of parameter combinations feasible and reasonable here is quite limited, which means large insecurity in the fits. However, these limitations are not severe, because we are not aiming here at quantitatively accurate description but rather to demonstrate the generic behaviour of the networks when control parameters are changed. The generic behaviour is adequately obtained despite the limited accuracy of the fits. The linear behaviour of κ_0 with the fitting function for the corresponding coefficients allows us to find the functional dependence of ΔK on β, needed for macroscopic description of the network.

The scaling function κ_0 allows us finally to obtain the maximum value K_0 and through it, the response function χ, which for a true thermodynamic system would correspond to the thermal capacity. The maximal value K_0 and response function χ are shown in Figure 3 at right. Their most important feature is that they increase with increasing value of β, which means that with higher β better resolving power (larger differences) is obtained. Also, the response function allows the interpretation that the larger the β, the more sensitive the value of divergence is to changes in structure of the network (i.e. with a large value of β better resolving power is possible). Another interesting notion is that the value of q mostly determines the behaviour of both, while values of λ affect them much less; the results are bunched according to the
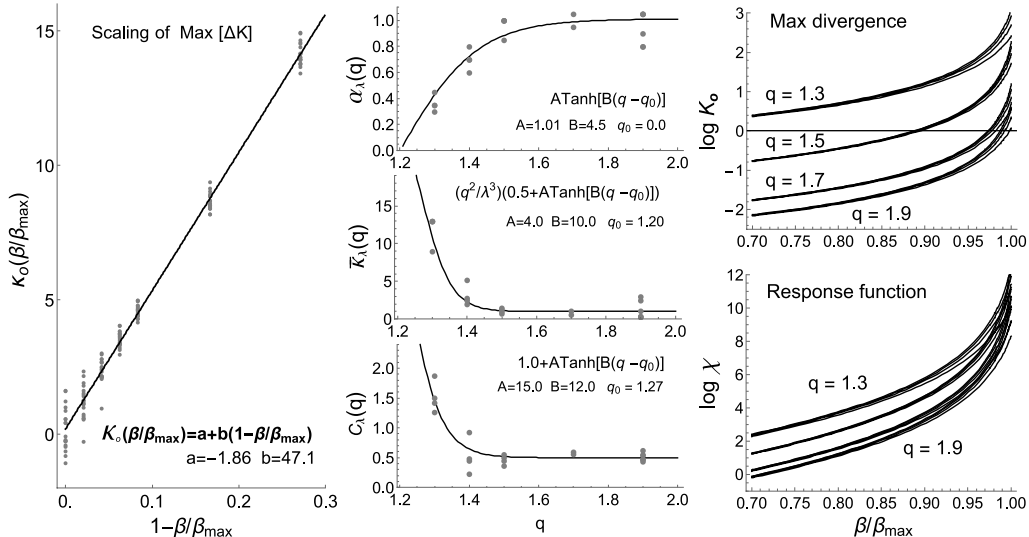
17

Fig. 3. The scaling functions for construction of free energy F_q. The dependence of maximum value κ_0 of ΔK is shown at the left panel, in scaled form and as function of $1 - \beta/\beta_{max}$. On the middle panel we show the fits to empirical data as used to construct α, $\bar{\kappa}$ and C_o for the semi-empirical functional form of free energy F_q. In these plots, the data-points are shown. On the right panel we show the semi-empirical construction of functional form of κ_0 for the maximum value of change in divergence (upper figure) and the corresponding function for the thermal response χ. In both figures, the dominant dependence is on parameter q and within each set of results, corresponding to given choice of q, a bunch of results corresponding to different choices of λ is visible. In each bunch, the highest value of $\lambda = 1.9$ corresponds to the topmost curve and in the curves below it correspond to the lower values $\lambda = 1.7$, 1.5 and 1.3 in that order.

value of q, and within these bunches, small differences originate from different choices for parameter q. Roughly, this means that differences in ΔK are more pronounced the smaller the value of q. This, of course, is expected in the case of heavy-tailed distributions since the smaller the values of q, the more sensitive the divergence is for low frequency values in distributions. Note, however, that if one wishes to retain the validity of Eq. (27) (and thus, also the validity of Eqs. (25) and (28)) the value of q for divergence cannot be chosen freely but must be same as for the q-adjacency kernel.

The present study leaves unanswered the question how the current model behaves when $q \to 1$. The theoretical argumentation presented here suggests that at that limit one should arrive at a description based on canonical ensemble (Gibbs' ensemble), von Neumann entropy and Kullback-Leibler divergence. However, for values $\lambda < 1.3$ reaching good computational stability is challenging and computations become quite unstable. Similarly, when of $q \to 2$ the computational results become very unstable. In both cases, reaching good

stability would require very extensive ensembles. The detailed reason for this is not known to us, but such behaviour is compatible with the notion that extrapolation from non-extensive thermodynamics of finite systems ($q > 1$) to the usual canonical extensive thermodynamics ($q = 1$) also requires the thermodynamic limit of infinite systems (compare with ref. [33]).

In summary, the main result contained in Figs. 2-3 is that the theoretically defined q-generalised free energy is a viable description of the macro-level state of the network changes when the modularity of the network is reduced. Within the thermodynamic-like interpretation, the change can be interpreted as lowering of the free energy due to relaxation of the modularity. In connection with this change, the probability of long paths increases. The effect is best seen for the adjacency kernels with low values of $q < 1.7$. The model also contains parameter β, which allows a different weighting of links, $\beta > 1$ corresponding to strong links, while $\beta < 1$ means making the links weaker, with the extreme case $\beta \to 0$ corresponding to totally disintegrating the network.

5 Conclusions

The practical motivation to conceptualise the properties of heavy-tailed networks from the macroscopic viewpoint derives from the notion that associative knowledge networks tend to be heavy-tailed, with broad tails that can be fitted with inverse power laws with powers in the range $1 < \lambda < 2$ and that in addition these networks are often highly modular [1–3]. However, real empirical networks are often quite small, and they cannot be easily or reliably characterised by pre-selected distributions of certain centrality like communicability or Katz centrality. Also, the similarity comparisons are often very awkward [1,2]. This prompts an attempt to use more holistic descriptions based on the network's density matrix, which provides all available information about the network [7,8]. Such description can be based on q-generalised adjacency kernels introduced here, which allows interpolation between known exponential and Neumann kernels.

The q-generalised adjacency kernel opens a door to the macro-level description of heavy-tailed networks, through the q-generalised Kullback-Leibler-Tsallis divergence. A parallel result has previously been obtained for closely similar q-generalised ensembles in the case on non-extensive statistics in general [33,45]. Here we have shown that analogously, the Kullback-Leibler-Tsallis divergence can be interpreted as a change of a q-generalised free energy of a modular heavy-tailed complex network, when modularity becomes relaxed. The finding that it is possible to define q-generalised free energy of heavy-tailed networks, and based on it, to obtain a thermodynamic-like interpretation, opens up interesting avenues to describe networks using thermodynamic-like concepts.

For example, it is shown that with the q-generalised free energy it is possible to derive the response of the network for changes of values of β, paralleling the thermal response if the thermodynamic interpretation is evoked. The response of the network with increasing value of β indicates that changes in the rigidity of the network (resilience) are smaller, the larger the values of β; the thermal capacity of the network increases when β increases.

The present study leaves many fundamental questions unanswered, most notably the question how the role of the adjacency matrix should be interpreted in terms of the Hamiltonian, and how far one can pursue the physical interpretations on the basis of structural, mathematical analogies. Nevertheless, we believe that the results provided here are promising steps towards a more general theory of heavy-tailed networks.

Appendix A: The q-communicabilities and q-divergence

This Appendix is a sequel to the main text and in it, we discuss communicability distributions. Starting from the q-generalized kernels, we define the q-generalized communicabilities. We then study their distributions arising from networks with heavy-tailed degree distributions, and use the q-divergences to differentiate between q-communicability distributions arising from modular and rewired networks.

The q-exponential adjacency kernel in Eq. 5 provides a basis for defining a q-generalised communicability (in brief, q-communicability) between nodes k and i as a row sum of q-adjacency kernel, in form

$$\Gamma_i(q, \beta) = Z^{-1} \sum_j \exp_q[\beta\, \mathbf{A}]_{ij} \qquad (A.1)$$

where $Z = \mathrm{Tr}\, \exp_q[\beta\mathbf{A}]$ is the normalisation factor. The q-generalised total communicability provides Estrada's total communicability in the limit $q \to 1$ and the Katz centrality is obtained in the limit $q \to 2$, corresponding to the exponential and Neumannn kernels, respectively.

For heavy-tailed networks, the q-communicability values are distributed according to an inverse power law, with inverse powers γ which are roughly in range from 2 to 3, but with a cut-off corresponding to largest observed value $\Gamma_{\max} = \mathrm{Max}[\{\Gamma_i\}]$. By rescaling the values of Γ_i by the maximum $\Gamma_{\max}$, and relabel $\Gamma_i/\Gamma_{\max} \to \Gamma_i \in [0,1]$, the new values can be represented in the form of the discrete distribution

$$p_i(\Gamma_i) = N^{-1} \left(1 + (\alpha - 1)\frac{1}{\epsilon}\Gamma_i\right)^{-1/(\alpha-1)}, \ \text{ with } \alpha \in\,]1,2]\,, \qquad (A.2)$$

where p_i is the probability of a given value $\Gamma_i \in [0,1]$ and N is the normalisation factor. The exponent α is related to inverse power γ as $\alpha = 1 + 1/\gamma$. In what follows, we use α in discussing the theoretical results while for simulation data γ is often more convenient. The parameter $\epsilon = \gamma/\Gamma_{\max} \ll 1$ has small values, but must always be different from zero. The distribution in Eq. (A.2) is q-exponential with exponent $q = \alpha$, where α is now used to refer to the q-index to avoid any confusion with index q in Eq. (5) associated with a q-adjacency kernel.

Results for power γ characterising the q-communicabilities in case $\Lambda = 1.5$ are shown in Figure A.1. The results are obtained for $q = \{1.05, 1.1, 1.2, \ldots, 2.0\}$ and with $\beta/\beta_{\max} = \{0.1, 0.2, 0.3, \ldots, 1.0\}$, where $\beta_{\max}$ is about 80 % of the inverse of the largest estimated eigenvalue of the adjacency matrix. The values of γ, and γ' in Fig. A.1 for fully rewired ($Q=0$) and original modular ($Q=0.86$)

21

networks, respectively, are obtained by least-square fits to log-log distributions. In fitting the coefficients, attention is paid to the tail of the distribution, and small values $\Gamma < 10^{-3}$ are allowed to deviate from the fits. The difference between powers γ and γ' is generally quite small, but detectable, as is seen in Fig. A.1 from the relative change $(\gamma' - \gamma)/\gamma$ (in the middle).

We can now utilise the q-divergence to compare q-communicability distributions with better resolution than provided by simply comparing the powers γ and γ'. With the parametrisation $\alpha = \alpha'(1 + \eta)$, where $\eta = (\alpha - \alpha')/\alpha' > 0$ is the relative difference, we can now obtain the q-divergence from Eq. (19) by replacing the density matrices with the discrete distributions and the traces of matrices by sums over the distributions, in the conventional way. A closed analytical form for the q-divergence is then obtained by replacing the discrete distribution (A.2) with the corresponding continuous distribution and evaluating the sum in Eq. (19) as a Riemann integral. These replacements are not trivial but can, however, be justified for q-exponentials [48]. The resulting expression for the q-divergence is then

$$K_\alpha(\eta, \epsilon) = \frac{1}{\alpha - 1} \left[\kappa_\alpha(\eta, \epsilon) \bar{K}_\alpha(\eta, \epsilon) \left(\epsilon^{\frac{-(2-\alpha)}{\alpha-1}} - (1 + \epsilon)^{\frac{-(2-\alpha)}{\alpha-1}} \right)^{\alpha-1} - 1 \right] \quad (A.3)$$

$$\kappa_\alpha(\eta, \epsilon) = \frac{(\alpha - 1)^{\alpha-1}(2 - \alpha(1 + \eta))^\alpha}{(2 - \alpha)^{\alpha-1}[2 - \alpha(1 + 2\eta)][\alpha(1 + \eta) - 1]^{\alpha-1}} \quad (A.4)$$

$$\bar{K}_\alpha(\eta, \epsilon) = \frac{\epsilon^{\frac{\alpha(1+2\eta)-2}{\alpha(1+\eta)-1}} - (1 + \epsilon)^{\frac{\alpha(1+2\eta)-2}{\alpha(1+\eta)-1}}}{\left(\epsilon^{1-\frac{1}{\alpha(1+\eta)-1}} - (1 + \epsilon)^{1-\frac{1}{\alpha(1+\eta)-1}} \right)^\alpha} \quad (A.5)$$

The parameter $\eta = (\alpha - \alpha')/\alpha' > 0$ is the relative difference of the q-indices, characterising the distribution, while the parameter ϵ is assumed to be the same for both distributions. Note that the parameter η attains always small values in the region of interest, where $\alpha = 1 + 1/\gamma$ with $\gamma \in]2, 3]$.

By using the above results with the values γ and γ' from Fig. A.1 we eventually obtain the q-divergence K_α shown in Fig. A.1 (at the right) as a function of the parameters q and β. We also used the analytical results in Eqs. (A.3)–(A.5) and in them, conversion $\eta = (\alpha' - \alpha)/\alpha$, where $\alpha = 1 + 1/\gamma$ and $\alpha' = 1 + \gamma'$. The main message of the result in Fig. A.1 is the following: From the values of K_α in Fig. A.1, we see that although the region close to the Katz centrality (the upper right corner of contour plot with $\beta > 0.8$ and $q > 1.7$) displays better resolution (i.e. large values of q-divergence), the region corresponding to Estrada's communicability at $q < 1.2$ and $\beta > 0.3$ also performs reasonably
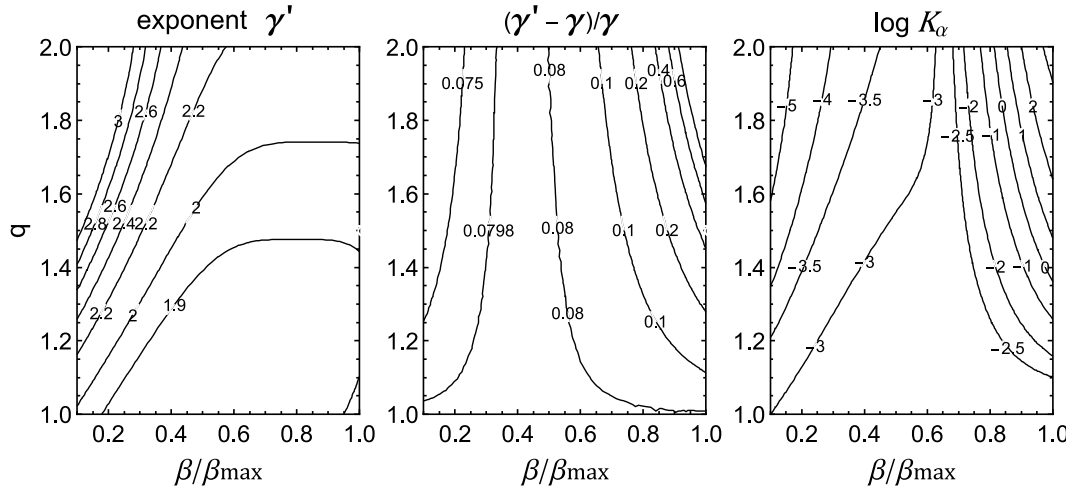
22

Fig. A.1. The contour plots of inverse powers γ' for modular network (at the left) and the relative difference $(\gamma' - \gamma)/\gamma$ (in the middle) with γ corresponding to a non-modular (rewired) distribution. The logarithm of q-divergence K_α is shown at the right. Results are shown for parameters $q \in]1, 2]$ and $\beta/\beta_{\max} \in [0.1, 1.0]$. Note that the valus of α and γ are related through $\alpha = 1 + 1/\gamma$. In reporting inverse powers we have used values of γ because the range of their variation is [2,3] while for α the corresponding range is narrower [1.33,1.5].

well in providing large enough divergences for good resolution. In this region the choice of β is not so crucial for the stability of the calculations as in the case of the region corresponding to the Katz centrality.

Finally, it should be mentioned that although the thermodynamic-like interpretation of q-divergence in Eqs. (A.3)–(A.5) is tempting, this is not viable now. The major complication in the thermodynamic interpretation arises now from the fact that communicability is a derived quantity which depends on parameter choices and does not refer to parameter independent and constitutive property of a network. Therefore, the Eq. (A.5) is better interpreted as divergence only. The utility of the analytical result in Eq. (A.5) is that it can be used to estimate the optimal region for the parameters with a compromise between a good resolution of divergence and computational simplicity.

The results demonstrate the practical utility and robustness of Estrada's communicability for applications where heavy-tailed distributions are involved. The results support Estrada's communicability as the most obvious choice, in many cases when a compromise between computational stability, ease, and sufficient accuracy is needed. This conclusion is of course already well-known [39]; the present study additionally shows, using the generalised q-adjacency kernels, that there is a smooth transition in the q-divergence between the regions corresponding to Estrada's communicability and the Katz centrality.

References

[1] I. T. Koponen, M. Nousiainen, University Students' Associative Knowledge of History of Science: Matthew Effect in Action?, Eur. J. Sci. Math. Educ. 6 (2018) 69–81.

[2] H. Lommi, I. T. Koponen, Network cartography of university students' knowledge landscapes about the history of science: landmarks and thematic communities, Appl. Netw. Sci. 4 (2019) 6.

[3] A. S. Morais, H. Olsson, L. J. Schooler, Mapping the Structure of Semantic memory, Cogn. Sci. 37 (2013) 125–145.

[4] R. Interdonato, M. Atzmueller, S. Gaito, R. Kanawati, C. Largeron, A. Sala, Feature-rich networks: going beyond complex network topologies, Appl. Netw. Sci. 4 (2019) 4.

[5] G. W. Thompson, C. T. Kello, Walking across Wikipedia: a scale-free network model of semantic memory retrieval, Front. Psychol. 5 (2014) 86.

[6] A. P. Masucci, A. Kalampokis, V. M. Equíluz, E. Hernández-García, Wikipedia Information Flow Analysis Reveals the Scale-Free Architecture of the Semantic Space, PLoS ONE 6 (2011) e17333.

[7] J. Biamonte, M. Faccin, M. De Domenico, Complex networks from classic to quantum, Comm. Phys. 2 (2019) 53.

[8] M. De Domenico, J. Biamonte, Spectral Entropies as Information-Theoretic Tools for Complex Network Comparison, Phys. Rev. X. 6 (2016) 041062.

[9] J. Kunegis, D. Fay, C. Bauckhage, Spectral evolution in dynamic networks, Knowl. Inf. Syst. 37 (2013) 1–36.

[10] M. Benzi, C. Klymko, On the Limiting Behavior of Parameter-Dependent Network Centrality Measures, SIAM Journal on Matrix Analysis and Applications, 36 (2015), pp. 686-706.

[11] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics, J. Stat. Phys. 52 (1988) 479–487.

[12] E. P. Borges, On a q-generalization of circular and hyperbolic functions, J. Phys. A: Math. Gen. 31 (1998) 5281–5288.

[13] T. Yamano, Some properties of q-logarithm and q-exponential functions in Tsallis statistics, Physica A. 305 (2002) 486–496.

[14] M. A. Ré, R. K. Azad, Generalization of Entropy Based Divergence Measures for Symbolic Sequence Analysis, PLoS ONE 9 (2014) e93532.

[15] J. P. Bagrow, E. M. Bollt, An information-theoretic, all-scales approach to comparing networks, Appl. Netw. Sci. 4 (2019) 45.

[16] G. Minello, L. Rossi, A. Torsello, Can a Quantum Walk Tell Which Is Which? A Study of Quantum Walk-Based Graph Similarity, Entropy 21 (2019) 328.

[17] F. Müller-Lennert, O. Dupuis, S. Szehr, M. Fehr, M. Tomamichel, On quantum Rényi entropies: A new generalization and some properties, J. Math. Phys. 54 (2013) 122203.

[18] G. A. Tsekouras, C. Tsallis, Generalized entropy arising from a distribution of q indices, Phys. Rev. E. 71 (2005) 046144.

[19] M. Gerlach, F. Font-Clos, E. G. Altmann, Similarity of Symbol Frequency Distributions with Heavy Tails, Phys. Rev. X 6, (2016) 021009.

[20] L. Dias, M. Gerlach, J. Scharloth, E. G. Altmann, Using text analysis to quantify the similarity and evolution of scientific disciplines. R. Soc. Open Sci. 5 (2018) 171545.

[21] E. G. Altmann, L. Dias, M. Gerlach, Generalized entropies and the similarity of texts, J. Stat. Mech. (2017) 014002.

[22] S. Abe, Nonadditive generalization of the quantum Kullback-Leibler divergence for measuring the degree of purification, Phys. Rev. A 68, (2003) 032302.

[23] S. Abe, Quantum q-divergence, Physica A 344 (2004) 359–365.

[24] A. F. T. Martins, N. A. Smith, E. P. Xing, Pedro M. Q. Aguiar, M. A. T. Figueiredo, Nonextensive Information Theoretic Kernels on Measures, J. Mach. Learn. Res. 10 (2009) 935-975.

[25] S. Furuichi, F.-C. Mitroi-Symeonidis, E. Symeonidis, On Some Properties of Tsallis Hypoentropies and Hypodivergences, Entropy 16 (2014) 5377-5399.

[26] I. T. Koponen, Modelling Students' Thematically Associated Knowledge: Networked Knowledge from Affinity Statistics, in: Complex Networks X. CompleNet 2019. Springer Proceedings in Complexity, Springer, Cham, 2019, pp. 123–134.

[27] A. D. Broido, A. Clauset, Scale-free networks are rare, Nat. Commun. 10 (2019) 1017.

[28] P. Holme, Rare and everywhere: Perspectives on scale-free networks, Nat. Commun. 10 (2019) 1016.

[29] M.E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113.

[30] V. D. P. Servedio, G. Caldarelli, P. Buttà, Vertex intrinsic fitness: How to produce arbitrary scale-free networks, Phys. Rev. E. 70 (2004) 056126.

[31] G. Caldarelli, A. Capocci, P. De Los Rios, M. A. Muñoz, Scale-Free Networks from Varying Vertex Intrinsic Fitness, Phys. Rev. Lett. 89 (2002) 258702.

[32] G. Csárdi, T. Nepusz, The Igraph software package for complex network research, InterJournal. Complex Systems (2006) 1695.

[33] S. Abe, A. K. Rajagopal, Validity of the Second Law in Nonextensive Quantum Thermodynamics, Phys. Rev. Lett. 91 (2003) 120601.

[34] E. Estrada, N. Hatano, M. Benzi, The physics of communicability in complex networks, Phys. Rep. 514 (2012) 89–119.

[35] E. Estrada, D. J. Higham, N. Hatano, Communicability betweenness in complex networks, Physica A. 388 (2009) 764–774.

[36] E. Estrada, The Structure of Complex Networks, Oxford University Press, Oxford, 2012.

[37] M. Dehmer, Information processing in complex networks: Graph entropy and information functionals. Appl. Math. Comput. 201 (2008) 82–94.

[38] M. Dehmer, A. Mowshowitz, A history of graph entropy measures, Inf. Sci. 181 (2011) 57–78.

[39] M. Benzi, E. Estrada, C. Klymko, Ranking hubs and authorities using matrix functions, Lin. Algb. Appl. 438 (2013) 2447–2474.

[40] L. Katz, A New Status Index Derived from Sociometric Analysis, Psychometrika 18 (1953) 39–43.

[41] K. J. Sharkey, A control analysis perspective on Katz centrality, Sci. Rep. 7 (2017) 17247.

[42] S. P. Borgatti, Centrality and network flow, Soc. Netw. 27 (2005) 55–71.

[43] R. Ghosh, K. Lerman, Parameterized centrality metric for network analysis, Phys. Rev. E. 83 (2011) 066118.

[44] M. Abramowitz, I. A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover, New York, 1972.

[45] S. Abe, Temperature of nonextensive systems: Tsallis entropy as Clausius entropy, Physica A. 368 (2006) 430–434.

[46] C. Tsallis, R. S. Mendes, A. R. Plastino, The role of constraints within generalized nonextensive statistics, Physica A. 261 (1998) 543–554.

[47] A. Fronczak, P. Fronczak, P., J. A. Hołyst, Fluctuation-dissipation relations in complex networks, Phys. Rev. E. 73 (2006) 016108.

[48] A. Plastino, M. C. Rocca, On the putative essential discreteness of q-generalized entropies, Physica A 488 (2017) 56–59.