

A unified data representation theory for network visualization, ordering and coarse-graining

István A. Kovács,^{1,2,*} Réka Mizsei,³ and Péter Csermely⁴

¹*Wigner Research Centre, Institute for Solid State Physics and Optics, H-1525 Budapest, P.O.Box 49, Hungary*

²*Institute of Theoretical Physics, Szeged University, H-6720 Szeged, Hungary*

³*Institute of Organic Chemistry, Research Centre for Natural Sciences,*

Hungarian Academy of Sciences, Pusztaszeri út 59-67, H-1025 Budapest, Hungary

⁴*Department of Medical Chemistry, Semmelweis University, H-1444 Budapest, P.O.Box 266, Hungary*

(Dated: October 1, 2014)

Representation of large data sets became a key question of many scientific disciplines in the last decade. Several approaches for network visualization, data ordering and coarse-graining accomplished this goal. However, there was no underlying theoretical framework linking these problems. Here we show an elegant, information theoretic data representation approach as a unified solution of network visualization, data ordering and coarse-graining. The optimal representation is the hardest to distinguish from the original data matrix, measured by the relative entropy. The representation of network nodes as probability distributions provides an efficient visualization method and, in one dimension, an ordering of network nodes and edges. Coarse-grained representations of the input network enable both efficient data compression and hierarchical visualization to achieve high quality representations of larger data sets. Our unified data representation theory will help the analysis of huge data sets in science, by revealing their large-scale structure in a comprehensible form.

PACS numbers: 89.75.Fb:Structures and organization in complex systems, 89.75.Hc:Networks and genealogical trees, 89.20.-a:Interdisciplinary applications of physics

Complex network^{1,2} representations are widely used in physical, biological and social systems, usually given by huge data matrices. Network data size grew to the extent, which is too large for direct comprehension and requires carefully chosen representations. One option to gain insight into the structure of complex systems is to order the matrix elements to reveal the concealed patterns, such as degree-correlations^{3,4} or community structure⁵⁻⁸. Currently, there is a diversity of matrix ordering schemes of different backgrounds, such as graph theoretic methods⁹, sparse matrix techniques¹⁰ and spectral decomposition algorithms¹¹. The coarse-graining or renormalization of networks¹²⁻¹⁷ also gained significant attention recently as an efficient tool to zoom out from the network, while reducing its size to a tolerable extent. A variety of heuristic coarse-graining techniques – also known as multi-scale approaches – emerged, leading to significant advances of network-related optimization problems^{18,19} and the understanding of network structure^{13,14}. The most essential tool of network comprehension is a faithful visualization of the network²⁰. Preceding more elaborate quantitative studies, it is capable of yielding an intuitive, direct qualitative understanding of complex systems. Although being of a primary importance, there is no general theory for network layout, leading to a multitude of graph drawing techniques. Among these, force-directed²¹ methods are probably the most popular visualization tools, which rely on physical metaphors. Graph layout aims to produce aesthetically appealing outputs, with many subjective aims to quantify – such as minimal overlaps between not related parts (e.g. minimal edge crossings in $d = 2$) –, while preserving the symmetries of the network. Altogether, the field of graph drawing became a meeting point of art, physics and computer science²².

Since the known approaches for the above problems generally lead to computationally expensive NP-hard problems²³, the practical implementations are necessarily restricted to advanced approximative heuristic algorithms. Moreover, there has been no successful attempt to incorporate network visualization, data ordering and coarse-graining into a common theoretical framework. Since information theory provides ideal tools to quantify the hidden structure in probabilistic data²⁴, its application to complex networks²⁵⁻²⁹ is a highly promising field. In this paper, our primary goal is to show an elegant, information theoretic representation theory for the unified solution of network visualization, data ordering and coarse-graining, when the input data has a probabilistic interpretation. Usually, in graph theory, the complex system is at the level of abstraction, where each node is a dimensionless object, connected by lines representing their relations, given by the input data. Instead, we study the case in which both the input matrix and the approximative representation is given in the form of a probability distribution. The best representation is selected by the criteria, that it is the hardest to distinguish from the input data. In information theory this is readily obtained by minimizing the relative entropy – also known as the Kullback-Leibler divergence³⁰ – as a quality function. In the following we show, that the visualization, ordering and coarse-graining of networks are intimately related to each other, being organic parts of a straightforward, unified representation theory.

General network representation theory

We consider the case, in which the input network is given by the symmetric, node-node co-occurrence A (adjacency) matrix having probabilistic entries $a_{ij} \geq 0$. If we start with the H edge-node co-occurrence (incidence) matrix instead, capable to describe hypergraphs as well, then $A \sim H^T H$ is simply given by the elements, $a_{ij} = \frac{1}{h_{**}} \sum_k h_{ki} h_{kj}$. Here and in the following asterisks indicate indices, for which we summed up. Throughout the paper we use the most general form of the input matrices, without assuming their normalization. Similarly, there is no need to normalize the information theoretic measures over \mathcal{A} , such as the S information content or the I mutual information given in the Methods section. The network is represented by a B co-occurrence matrix and a natural way to quantify the quality of the representation is to use the relative entropy (D). The relative entropy measures the extra description length, when B is used to encode the data described by the original matrix, A ,

$$D(A||B) = \sum_{ij} a_{ij} \ln \frac{a_{ij} b_{**}}{b_{ij} a_{**}} \geq 0. \quad (1)$$

Although $D(A||B)$ is not a metric and not symmetric in A and B , it is an appropriate and widely applied measure of statistical remoteness³¹, quantifying the distinguishability of B from A . Thus, the highest quality representation is achieved, when the relative entropy approaches 0, and our general goal is to obtain a B^* representation as

$$B^* = \operatorname{argmin}_B D(A||B). \quad (2)$$

Since $D(A||B) = H(A, B) - S(A)$, where $H(A, B)$ is the (unnormalized) cross-entropy, we could equivalently minimize the cross-entropy for B . Although the minimization of $D(A||B)$ appears in the *minimal discrimination information*, also known as the *minimum cross-entropy* (MinxEnt) approach by Kullback³², there the goal is the opposite of ours, namely to find the optimal 'real' distribution, A , while the 'approximate' distribution, B , is kept fixed. In this sense, our optimization is an inverse MinxEnt problem³³. This kind of optimization appears also as a refinement step to improve the importance sampling in Monte Carlo methods (for highly restricted A -s), under the name of *cross-entropy method*³⁴. In order to avoid confusion and emphasize the differences, we only use the term of relative entropy in the following.

Although $D(A||B) \geq 0$ can be arbitrarily large, there is always available a *trivial* representation by the uncorrelated, *product state*, B^0 matrix given by the elements $b_{ij}^0 = \frac{1}{a_{**}} a_{i*} a_{*j}$. This way $D_0 \equiv D(A||B^0) = I \leq S$ is the I mutual information, thus the optimized value of D can be always normalized with I , or alternatively as $\eta \equiv D/S \leq 1$. Here, η gives the ratio of the needed extra description length to the optimal description length of the system. The optimization of relative entropy is local in the sense, that the global optimum of a network comprising independent subnetworks is also locally optimal for each subnetwork.

The finiteness of D_0 ensures, that if i and j are connected in the original network ($a_{ij} > 0$), then they are guaranteed to be connected in a meaningful representation as well, enforcing $b_{ij} > 0$, since otherwise D would diverge. In the opposite case, when we have a connection in the representation, without a corresponding edge in the original graph ($b_{ij} > 0$ while $a_{ij} = 0$), b_{ij} does not appear directly in D , only globally, through the b_{**} normalization. Nevertheless, the B matrix of the optimal representation (where D is small) is *close* to A , since the total variation of the normalized distributions is bounded by D ³⁵

$$D(A||B) \geq \frac{a_{**}}{2 \ln 2} \sum_{i,j} \left(\frac{a_{ij}}{a_{**}} - \frac{b_{ij}}{b_{**}} \right)^2. \quad (3)$$

Thus, in the optimal representation of the network all the connected network elements are connected, while having only a strongly suppressed amount of false positive connections.

Network visualization and data ordering

Since force-directed layout schemes²¹ have an energy or quality function, optimized by efficient techniques borrowed from many-body physics³⁶ and computer science³⁷, graph layout could be in principle serve as a quantitative tool. However, these approaches inherently struggle with an information shortage problem, since the edge weights only provide half the needed data to initialize these techniques. For instance, for the initialization of the Fruchterman-Reingold³⁸ (Kamada-Kawai³⁹) method we need to set both the strength of an attractive force (optimal distance) and a repulsive force (spring constant) between the nodes in order to have a balanced system. Due to the lack of sufficient information, such graph layout techniques become somewhat ill-defined and additional subjective considerations are

needed to double the information encoded in the input data, traditionally by a nonlinear transformation of the attractive force parameters onto the parameters of the repulsive force.

While in usual graph layout schemes the graph nodes are represented by points (without spatial extension) in a d -dimensional background space – connected by (straight, curved or more elaborated) lines – in our approach the nodes are extended objects, namely probability distributions (ρ) over the background space. The role of edges is played by the overlaps of the node distributions. Importantly, in our representation the shape of nodes encodes just that additional set of information, which has been lost and then arbitrarily re-introduced in the above mentioned network visualization methods. In the following we consider the simple case of Gaussian distributions – having a width of σ , and norm $h = \int \rho$, see equation (6) of the Methods section –, but we have also tested the non-differentiable case of a homogeneous distribution in a spherical region of radius σ . For a given graphical representation the B co-occurrence matrix is built up from the overlaps of the distributions ρ_i and ρ_j - analogously to the construction of A from H - as $b_{ij} = \frac{1}{R} \int dx^d \rho_i(x) \rho_j(x)$, where $R = \sum_k \int dx^d \rho_k(x)$ is an (irrelevant) normalization factor.

The trivial data representation of B^0 can be obtained by an initialization, where all the nodes are at the same position, with the same distribution function (apart from a varying $h_i \propto a_{i*}$ normalization to ensure the proper statistical weight of the rows). This way, initially $D_0 = I$ is the mutual information of the input matrix, irrespectively from the chosen distribution function. Here we note, that the b_{ii} diagonal self-overlaps cannot change during the repositioning of the distributions in the layout, only due to changing the shape parameters of the distributions, namely the σ_i widths or h_i normalizations. The numerical optimization can be in general, straightforwardly carried out by a usual simulated annealing scheme starting with an initialization of B^0 . Alternatively, in the differentiable case we can also use a Newton-Raphson iteration as in the Kamada-Kawai method³⁹ (for details see the Methods section). In terms of the layout, the finiteness of D_0 ensures that the connected nodes overlap in the layout as well, even for distributions having a finite support. Moreover, independent parts of the network (nodes or sets of nodes without connections between them) tend to be apart from each other in the layout. Additionally, if two rows (and columns) of the input matrix are proportional to each other, then it is optimal to represent them with the same distribution function in the layout, as though the two rows were fused together.

When using Gaussian distributions, our visualization method can be conveniently interpreted as a force-directed method. If the normalized overlap, b_{ij}/b_{**} , is smaller at a given edge than the normalized edge weight, a_{ij}/a_{**} , then it leads to an attractive force, while the opposite case induces a repulsive force. For Gaussian distributions all the nodes overlap in the representations, leading typically to $D > 0$ in the optimal representation. However, for distributions with a finite support, such as the above mentioned homogeneous spheres, perfect layouts with $D = 0$ can be achieved even for sparse graphs. In $d = 2$ this concept is reminiscent to the celebrated concept of planarity⁴⁰. However, our concept can be applied in any dimensions. Furthermore, any network of $I = 0$ (e.g. a fully connected graph) is perfectly represented in any dimensions by B^0 , that is by simply putting all the nodes at the same position. Here we note, that the concept of cross-entropy have already appeared in the field of graph drawing, such as in the methods of Refs. 41,42. Besides others, the most important difference between these methods and ours is, that in our case the relative entropy (or cross entropy) is calculated over $N \times N$ -point distributions for N nodes, while in the related papers^{41,42} only 2-point distributions of the form $\{a_{ij}, 1 - a_{ij}\}$ are considered.

Our method is illustrated in Fig 1. on the Zachary karate club network⁴³, which became a corner stone of graph algorithm testing. It is a weighted social network of friendships between $N_0 = 34$ members of a karate club at a US university, which fell apart after a while into two communities indicated by different colors in Fig. 1. Our network layout technique works in any dimensions, as illustrated in $d = 1, 2$ and 3. In each case the communities are clearly recovered and, as expected, the quality of layout becomes better as the dimensionality of the embedding space increases.

Nevertheless, the one dimensional case deserves special attention, since it serves as an ordering of the elements as well (after resolving possible degenerations with small perturbations), as illustrated in Fig 1.e. Although our network layout works only for symmetric co-occurrence matrices, the ordering can be extended for hypergraphs with asymmetric H matrices as well, since the orderings of the two adjacency matrices HH^T and H^TH readily yield an ordering for the rows and columns of the matrix, H .

When applying a local scheme (e.g. simulated annealing) for the optimization of the representations, we generally run into computationally hard situations. These correspond to local minima, in which the layout can not be improved by single node updates, since whole parts of the network should be updated (rescaled, rotated or moved over each other), instead. Being a general difficulty in many optimization problems, it was expected to be insurmountable also in our approach. In the following we show, that the relative entropy based coarse-graining scheme – given in the next section – can, in practice, efficiently help us trough these difficulties in polynomial time.

Coarse-graining of networks

Since it is generally expected to be an NP-hard problem to find the optimal simplified, coarse-grained description of a network at a given scale, we have to rely on approximate heuristics having a reasonable run-time. In the following we use a local coarse-graining approach, where in each step a pair of rows (and columns) is replaced by coarse-grained rows, giving the best approximative new network G' in terms of $D(G||G')$, where G is the H or A matrix of the initial network. Although the applied formulas and coarse-graining steps are different, the idea of pairwise, information theoretic coarse-graining appeared recently also in the method of Ref. 28 to detect the presence of meso-scale structures in complex networks. In our method, for $G = H$ the coarse-graining step means, that instead of the original k and l rows, we use two new rows, being proportional to each other, while the $h'_{k*} = h_{k*}$ and $h'_{l*} = h_{l*}$ probabilities are kept fixed

$$h'_{ki} = h_{k*} \frac{h_{ki} + h_{li}}{h_{k*} + h_{l*}}, \quad h'_{li} = h_{l*} \frac{h_{ki} + h_{li}}{h_{k*} + h_{l*}}. \quad (4)$$

This way the rows involved are replaced by their normalized average. Figuratively, this way 'bonds' are formed between the nodes in each step with a given $D(G||G')$ 'scale'. As in the case of graph layout, if two rows (or columns) are proportional to each other, they can be fused together, since their coarse-graining leads to $D = 0$. Consequently, we can alternatively think of the coarse-graining step as a fusion or grouping of the rows involved. For an illustration of the fused data matrices see the lower panels of Fig. 2 a-d. We note, that proportional rows (or columns) can be generally fused together also initially, in the input data, as a prefiltering, before starting to find an optimal representation.

Since in a coarse-graining step the matrix is modified, for the remaining steps the D difference should be updated if at least one member of the pair is neighbour of the fused elements. For n elements this takes $\mathcal{O}(n^3)$ time until the whole network is fused together. When $G = A$, the coarse-graining step is carried out simultaneously and identically for the rows and columns. The optimal coarse-graining is illustrated in Fig. 1.f for the Zachary karate club network. The heights in the dendrogram indicate the D values of the representations when the fusion step happens. Although it seems to be somewhat tedious in a later step to measure D always from the original input network for a given (k, l) pair, there is a simple rule to calculate this from the actually existing coarse-grained data alone. If D_k and D_l are the values when the rows (and columns) k and l were formed via fusion (being zero initially), then from the apparent D' value – measuring the formation of a bond directly from the coarse-grained rows k and l – we get

$$D = D_k + D_l + D'. \quad (5)$$

Since the normalization of the rows and columns is preserved during coarse-graining, finally we arrive at $D = I = D_0$. This nicely coincides with the above proposed initialization step of the network layout approach.

Hierarchical layout

Although the introduced coarse-graining scheme may be of significant interest whenever probabilistic matrices appear, here we are primarily interested in its application for network layout, to obtain a hierarchical visualization^{44–49}. Our bottom-up coarse-graining results can be readily incorporated into the network layout scheme in a top-down way by initially (at high temperature) starting with one node (comprising the whole system), and successively undoing the fusion steps (cutting bonds) until the original system is recovered. Between each such extension step the layout can be optimized as before.

We have found, that this hierarchical layout scheme produces significantly improved layouts compared to a local optimization, such as a simple simulated annealing. By incorporating the coarse-graining in a top-down approach, we first arrange the position of the large-scale parts of the network, and refine the picture in later steps only. The refinement steps happen, when the position and extension of the large-scale parts have already been sufficiently optimized. After such a refinement step, the nodes – moved together so far – are treated separately. At a given scale (having $N \leq N_0$ nodes), the D value of the coarse-graining provides a lower bound for the D value of the obtainable layout. Our hierarchical visualization approach is illustrated in Fig. 2. with snapshots of the layout and the coarse-grained representation matrix of the Zachary karate club network⁴³ at $N = 5, 15, 25$ and 34 . As an illustration on a larger and more challenging network, in Fig. 3. we show the result of the hierarchical visualization on the giant component of the weighted human disease network⁵⁰. In this network we have $N_0 = 516$ nodes, representing diseases, connected by mutually associated genes. The colors indicate the known disease groups. In the numerical optimization for this network we primarily focused on the positioning of the nodes, thus the optimization for the widths and normalizations was only turned on as a fine-tuning after an initial layout was obtained.

Being an efficient approximation scheme to speed up the optimization process, it is pointless to spend huge computational efforts to carry out the coarse-graining steps in a large network precisely. It leads to a faster and reasonable approximation to leave out (some of) the recalculation steps. As a quadratic, $\mathcal{O}(N^2)$, approach we can fuse together all the nodes with their best match, according to the initially obtained D values. This procedure uses $\mathcal{O}(\log N)$ global fusion steps. An even faster approach is to leave out the recalculation steps completely. In this simplified scheme, the coarse-graining happens on the minimal spanning tree of the original matrix of pairwise calculated D values. While for smaller networks ($N \sim 10^3$) the full approach is feasible, for larger networks we suggest to use the faster, approximative methods.

Discussion

In this paper we have shown, that a straightforward information theoretic idea provides an elegant, unified solution for such long-standing problems, as matrix ordering, network visualization and data coarse-graining. First, we demonstrated that the minimization of relative information yields a novel visualization technique, while representing the A input matrix by the B co-occurrence matrix of extended distributions, embedded in a d -dimensional space. As another application of the same approach we obtained a hierarchical coarse-graining scheme, when the input matrix is represented by its subsequently coarse-grained versions. Since these applications are two sides of the same representation theory, they turned out to be superbly compatible, leading to an even more powerful hierarchical visualization technique. Although we have focused on the visualization in d -dimensional continuous space, the representation theory can be applied more generally, incorporating also the case of discrete embedding spaces. A possible future application is the optimal embedding of a (sub)graph into another graph. Our relative entropy-based visualization with e.g. Gaussians can be naturally interpreted as a force-directed method. Traditional force directed methods induced huge efforts on the computational side to achieve scalable algorithms applicable for the large data sets in real life. Here we naturally can not and do not wish to compete with such advanced techniques, but believe that our approach can be a good starting point for further scalable implementations. As shown, network visualization is already interesting in one dimension yielding an ordering for the elements of the network. Our approach is formulated directly for the adjacency or incidence matrix of weighted networks, incorporating also the cases of bipartite graphs and hypergraphs. Since in this paper our primary intention was merely to demonstrate our theoretical framework, the more involved analysis of further interesting networks will be the subject of forthcoming articles. Due to our network representation theory we hope to get closer to discover knowledge from the huge data matrices in science.

Methods

The (unnormalized) Shannon entropy of \mathcal{A} , expressing the amount of information in the system is given by $S = -\sum_{ij} a_{ij} \ln \frac{a_{ij}}{a_{**}}$, while the \mathcal{I} mutual information between the rows and columns of A is given by $\mathcal{I} = \sum_{ij} a_{ij} \ln \frac{a_{ij} a_{**}}{a_{i*}^2}$. The parametrization of the Gaussian distributions used in the layout is the following in d -dimensions

$$\rho(\{x_i^0\}, \sigma, n) = \frac{n}{\sqrt{(2\pi)^d}} \exp\left(-\frac{\sum_{i=1}^d (x_i - x_i^0)^2}{2\sigma^2}\right). \quad (6)$$

Here we note, that in practice, special care may be needed for the a_{ii} diagonal elements of A , describing the probability of the co-occurrence of an element with itself. If the nodes represent individual entities, rather than some properties or groups, then such self co-occurrences are impossible leading to $a_{ii} \equiv 0$, which should be included in the representation scheme as well, by requiring $b_{ii} \equiv 0$. While the solution of this case is rather straightforward, for the sake of simplicity we omitted its detailed description here.

For the numerical optimization of the network layout, we have implemented a simple, general purpose simulated annealing scheme, and for Gaussian distributions a much faster Newton-Raphson update, applied also in the Kamada-Kawai method³⁹ and in Refs. 41,42. In practice, we used a separate Newton-Raphson step for the d coordinates and for the σ_i widths and h_i normalizations. In each step, the node with the largest gradient amplitude ($\|J\|$) is updated in the direction and with a parameter step size, obtained by the second derivative matrix, \mathcal{F} as $-\mathcal{F}^{-1}J$. Since \mathcal{F} is not always positive definite, special care is needed when the relative entropy increases in such a step. In such a case, a sufficiently small step size is applied in the direction of the gradient vector, instead. This way our technique has the same computational complexity as the widely applied Kamada-Kawai method³⁹.

* Electronic address: kovacs.istvan@wigner.mta.hu

¹ Newman, M. E. J. *Networks-An Introduction*. Oxford Univ. Press, (2010).

² Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics* **74**, 47-97 (2002).

³ Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).

⁴ Reshef, D. N. et al. Detecting Novel Associations in Large Data Sets. *Science* **334**, 1518-1524 (2011).

- ⁵ Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA* **99**, 7821-7826 (2002).
- ⁶ Newman, M. E. J. Communities, modules and large-scale structure in networks. *Nature Physics* **8**, 25-31 (2012).
- ⁷ Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75-174 (2010).
- ⁸ Kovács, I. A., Palotai, R., Szalay, M. S. & Csermely, P. Community landscapes: an integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. *PLoS ONE* **5**, (9):e12528 (2010).
- ⁹ King, I. P. An automatic reordering scheme for simultaneous equations derived from network analysis. *Int. J. Numer. Methods*, **2**, 523-533 (1970).
- ¹⁰ George A. & Liu, J. W.-H. Computer Solution of Large Sparse Positive Definite Systems. Prentice-Hall Series in Computational Mathematics. Prentice-Hall, Inc, Englewood Cliffs, NJ, (1981).
- ¹¹ West, D. B. Introduction to Graph Theory. Prentice-Hall, Inc. (1996).
- ¹² Song, C., Havlin, S. & Makse, H. A. Self-similarity of complex networks. *Nature* **433**, 392-395 (2005).
- ¹³ Radicchi, F., Ramasco, J. J., Barrat, A. & Fortunato, S. Complex Networks Renormalization: Flows and Fixed Points. *Phys. Rev. Lett.* **101**, 148701 (2008).
- ¹⁴ Rozenfeld, H. D., Song, C. & A. Makse, H. A. Small-World to Fractal Transition in Complex Networks: A Renormalization Group Approach. *Phys. Rev. Lett.* **104**, 025701 (2010).
- ¹⁵ Gfeller, D. & De Los Rios, P. Spectral Coarse Graining of Complex Networks. *Phys. Rev. Lett.* **99**, 038701 (2007).
- ¹⁶ Sales-Pardo, M., Guimera, R., Moreira, A. A. & Amaral L. A. N. Extracting the hierarchical organization of complex systems. *Proc. Natl. Acad. Sci. USA* **104**, 15224-15229 (2007).
- ¹⁷ Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A.-L. Hierarchical Organization of Modularity in Metabolic Networks. *Science* **297**, 1551-1555 (2002).
- ¹⁸ Walshaw, C. A multilevel approach to the travelling salesman problem. *Oper. Res.*, **50**, 862-877 (2002).
- ¹⁹ Walshaw, C. Multilevel refinement for combinatorial optimisation problems. *Annals of Operations Research* **131**, 325-372 (2004).
- ²⁰ Di Battista, G., Eades, P., Tamassia, R. & Tollis, I.G. Algorithms for the visualization of Graphs. Prentice-Hall, (1999).
- ²¹ Kobourov, S. G. Spring Embedders and Force-Directed Graph Drawing Algorithms. arXiv:1201.3011 (2012).
- ²² North, S. C. (Ed.) Graph Drawing, Symposium on Graph Drawing GD'96, Berkeley, California, USA, September 18 - 20, 1996, Proceedings, Springer Lecture Notes in Computer Science, ISBN 3-540-62495-3 (1997).
- ²³ Garey, M. R. & Johnson, D. S. Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman and Co., (1979).
- ²⁴ Kinney, J. B. & Atwal, G. S. Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. USA* **111** no. 9, 3354-3359 (2014).
- ²⁵ Slonim, N., Atwal, G. S., Tkačik, G. & Bialek, W. Information-based clustering. *Proc. Natl. Acad. Sci. USA* **102** no. 51 18297-18302 (2005).
- ²⁶ Rosvall, M. & Bergstrom, C. T. An information-theoretic framework for resolving community structure in complex networks. *Proc. Natl. Acad. Sci. USA* **104**, 7327-7331 (2007).
- ²⁷ Rosvall, M., Axelsson, D. & Bergstrom, C. T. The map equation. *Eur. Phys. J. Special Topics* **178**, 13 (2009).
- ²⁸ Zanin, M., Sousa, P. A. & Menasalvas, E. Information content: assessing meso-scale structures in complex networks. *Europhys. Lett.* **106** (3), 30001 (2014).
- ²⁹ Allen, B., Stacey, B. C. & Bar-Yam, Y. An Information-Theoretic Formalism for Multiscale Structure in Complex Systems. arXiv:1409.4708 (2014).
- ³⁰ Kullback, S. & Leibler, R. A. On Information and Sufficiency. *Annals of Mathematical Statistics* **22**, 79-86 (1951).
- ³¹ Grünwald, P. D., The Minimum Description Length Principle, MIT Press. (2007).
- ³² Kullback, S. Information Theory and Statistics, John Wiley: New York, NY, USA (1959).
- ³³ Kapur, J. N. & Kesavan, H. K., The Inverse MaxEnt and MinxEnt Principles and their Applications, in Maximum Entropy and Bayesian Methods, Fundamental Theories in Physics, Springer Netherlands, **39**, 433-450 (1990).
- ³⁴ Rubinstein, R. Y., The cross-entropy method for combinatorial and continuous optimization. *Method. Comput. Appl. Probab.* **1**, 127-190 (1999).
- ³⁵ See, e.g., Cover, Th. M. & Thomas, J. A. Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing) (Lemma 12.6.1, pp. 300-301) (1991).
- ³⁶ Barnes, J. & Hut, P. A hierarchical O(NlogN) force-calculation algorithm. *Nature*, **324**, 446-449 (1986).
- ³⁷ Gansner, E. R., Koren, Y. & North, S. Graph Drawing by Stress Majorization. Proceedings of 12th Int. Symp. Graph Drawing, Lecture Notes in Computer Science **3383**, Springer-Verlag, 239-250 (2004).
- ³⁸ Fruchterman, T. M. & Reingold, E. M. Graph Drawing by Force-Directed Placement, Software: Practice & Experience (Wiley) **21**, 1129-1164 (1991).
- ³⁹ Kamada, T. & Kawai, S. An algorithm for drawing general undirected graphs. *Information Processing Letters* (Elsevier) **31**, 7-15 (1989).
- ⁴⁰ Hopcroft, J. & Tarjan, R. E. Efficient planarity testing. *Journal of the Association for Computing Machinery* **21**, 549-568 (1974).
- ⁴¹ Yamada, T., Saito, K. & Ueda, N. Cross-entropy directed embedding of network data, In Proceedings of the 20th International Conference on Machine Learning (ICML2003), p. 832-839 (2003).
- ⁴² Estévez, P. A., Figueroa, C. J. & Saito, K. Cross-entropy embedding of high-dimensional data using the neural gas model. *Neural Networks*, **18**, 727-737 (2005).
- ⁴³ Zachary, W. W. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research* **33**,

- 452-473 (1977).
- ⁴⁴ Hu, Y. F. Efficient and high quality force-directed graph drawing. *The Mathematica Journal*, **10**, 37-71 (2006).
- ⁴⁵ Szalay-Bekő, M., Palotai, R., Szappanos, B., Kovács, I. A., Papp, B. & Csermely P., ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality. *Bioinformatics* **28**, 2202-2204 (2012).
- ⁴⁶ Six, J. M., & Tollis, I. G. Effective Graph Visualization Via Node Grouping, in Software Visualization, The Springer International Series in Engineering and Computer Science **734**, 413-437 (2003).
- ⁴⁷ Goodrich, M. T., Gajer, P. & Kobourov, S. G. A fast multi-dimensional algorithm for drawing large graphs. *Lecture Notes In Computer Science*, **1984**, 211-221 (2000).
- ⁴⁸ Harel, D. & Koren, Y. A fast multi-scale method for drawing large graphs. *J. Graph Algorithms and Applications*, **6**, 179-202 (2002).
- ⁴⁹ Walshaw, C. A multilevel algorithm for force-directed graph drawing. *J. Graph Algorithms Appl.*, **7**, 253-285 (2003).
- ⁵⁰ Goh, K.-I. et al. The human disease network. *Proc. Natl. Acad. Sci. USA* **104**, 8685-8690 (2007). The network data can be downloaded from the following site: http://www.barabasilab.com/pubs/CCNR-ALB_Publications/200705-14_PNAS-HumanDisease/Suppl/index.htm

Acknowledgements

We are grateful to the members of the LINK-group (www.linkgroup.hu) for useful discussions. This work was supported by the Hungarian National Research Fund under grant Nos. OTKA K109577 and K83314. The research of IAK was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP 4.2.4. A/2-11-1-2012-0001 'National Excellence Program'.

Author contributions

IAK and RM conceived the research and ran the numerical simulations. IAK devised and implemented the applied algorithms. IAK and PCs wrote the paper.

Additional information

The authors declare no competing financial interests. Correspondence should be addressed to I.A.K..

Figures

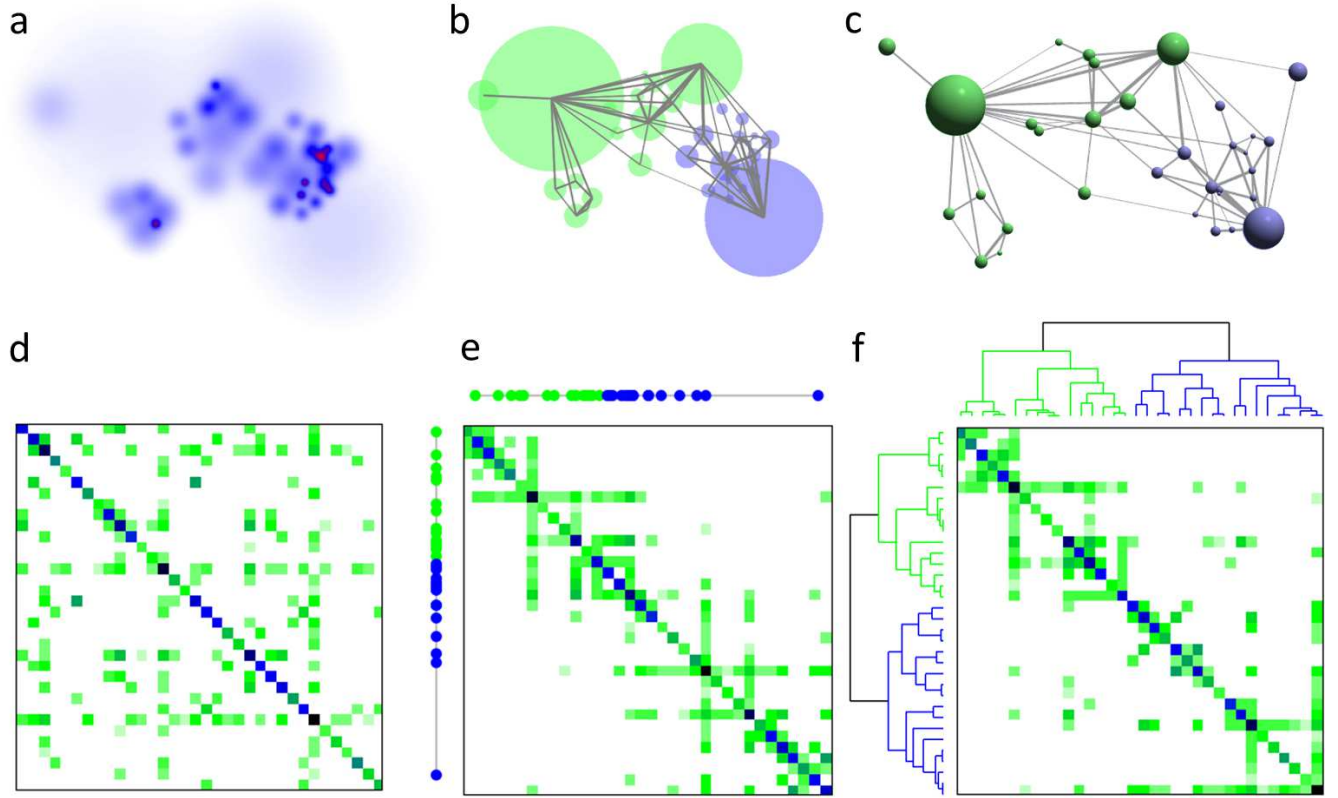


FIG. 1: Illustration of the power of our unified representation theory on the Zachary karate club network⁴³. The optimal layout ($\eta = 2.1\%$) in terms of $d = 2$ dimensional Gaussians is shown by a density plot in **a** and by circles of radiuses σ_i in **b**. **c** the best layout is obtained in $d = 3$ ($\eta = 1.7\%$), where the radiuses of the spheres are chosen to be proportional to σ_i . **d** the original data matrix of the network with an arbitrary ordering. **e** the $d = 1$ layout ($\eta = 4.5\%$) yields an optimal ordering of the original data matrix of the network. **f** the optimal coarse-graining of the data matrix yields a tool to zoom out from the network in accordance with the underlying community structure. We stress, that the coarse-graining itself does not yield a unique ordering of the nodes, therefore an arbitrary chosen compatible ordering is shown in this panel.

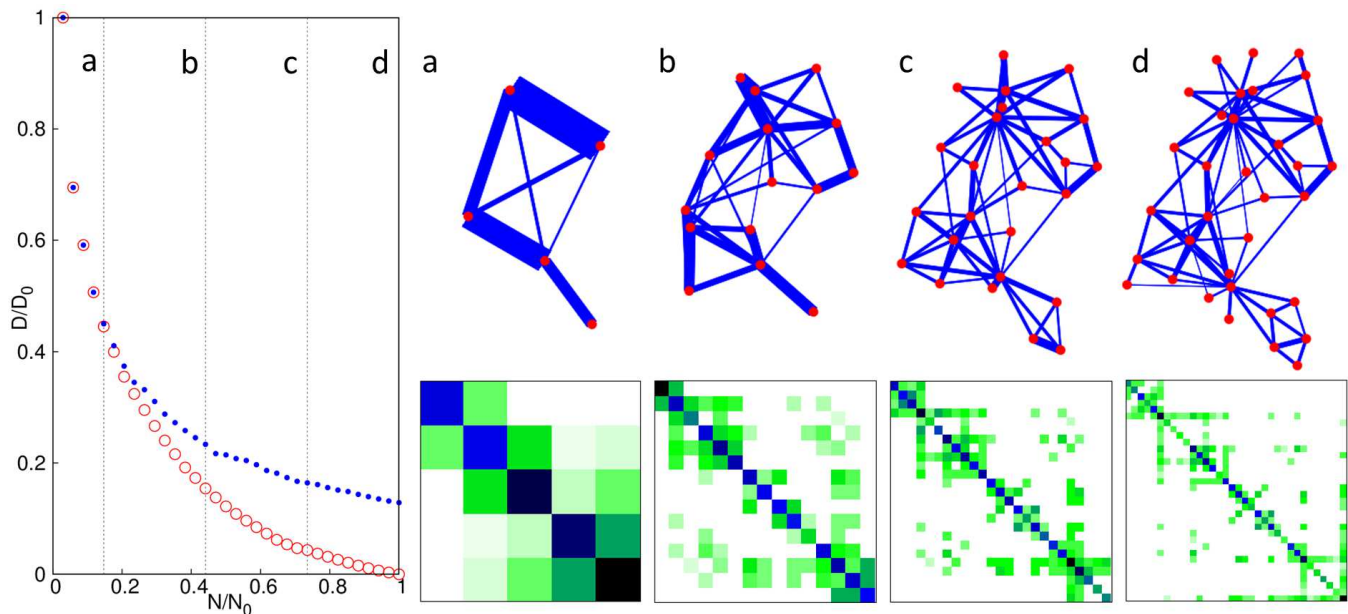


FIG. 2: **Illustration of our hierarchical visualization technique on the Zachary karate club network⁴³.** In our hierarchical visualization technique the coarse-graining procedure guides the optimization for the layout in a top-down way. As the N number of nodes increases, the relative entropy of both the coarse-grained description (red, \circ) and the layout (blue, \bullet) decreases. The panels **a-d** show snapshots of the optimal layout and the corresponding coarse-grained input matrix at the level of $N = 5, 15, 25$ and 34 nodes, respectively. For simplicity, here the h_i normalization of each distribution is kept fixed to be $\propto a_{i*}$ during the process, leading finally to $\eta = 4.4\%$.

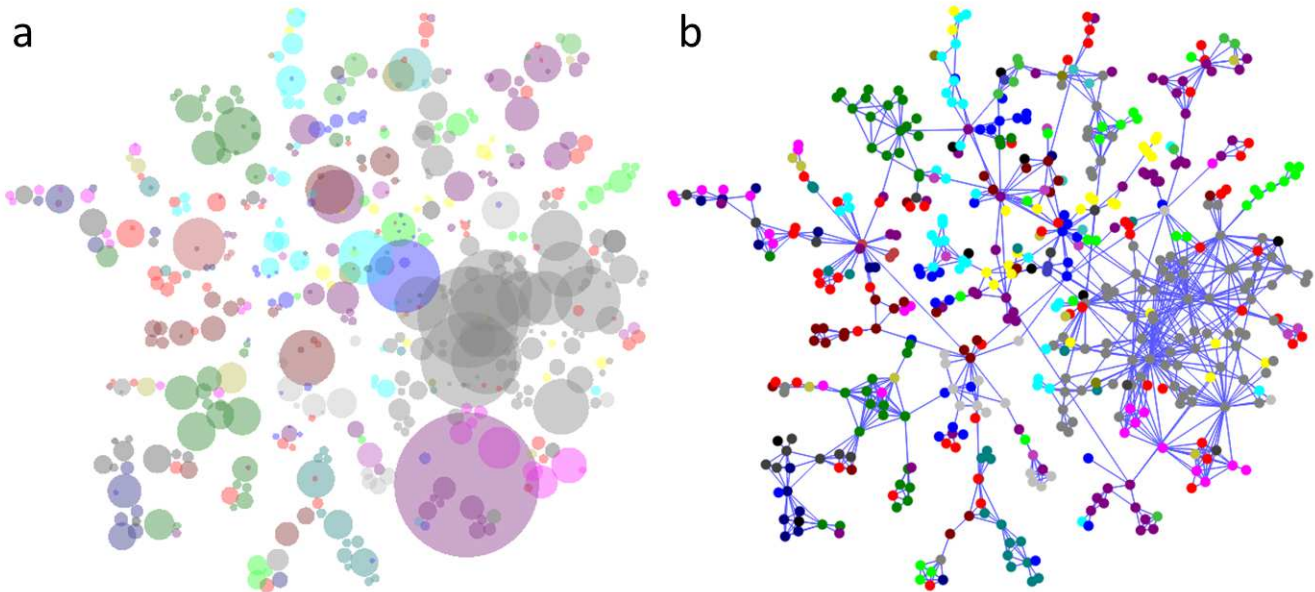


FIG. 3: **Visualization of the human diseaseome.** The obtained best layout ($\eta = 3.1\%$) by our hierarchical visualization technique of the human diseaseome is shown by circles of radiuses σ_i in **a** and by a traditional graph in **b**. The nodes represent diseases, colored according to known disease categories⁵⁰.