On the role of Symmetry in the Community-Finding Problem in Complex Networks: Statistical Mechanics of clustering

Miguel A. Santos
*Centre for Computational Biology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, M5G 1X8**
(Dated: April 15, 2019)

This text represents the personal notes of the author on a work in progress. We describe the role of symmetry in the problem of finding communities in a complex network. We outline a possible matrix representation of partitions that carries a geometric interpretation as rotations in a space of N dimensions. We discuss the possibility of re-defining Zurek's algorithmic distance measure such that it takes into account the *a priori* information defining the clustering problem. This leads us to outline a general theoretical approach for identifying clusters among a set of elements in a way that is analogous to the MaxEntropy approach of Statistical Mechanics. The graph of pairwise affinities between elements is the basic input. Fixing the appropriate averages should determine the corresponding optimal clustering.

Contents

I. INTRODUCTION

A. Overview

Since the seminal works of Strogatz and Watts [1], Barabási and Albert [2] and that of Girvan and Newman [3], complex networks in general, but in particular the community finding problem is being devoted a constantly increasing interested within the fields of Physics and Mathematics [4–6]. However, many other fields sparked the initial interest in these topics, albeit disguised under different names. Indeed, clustering a set of objects into disjoint classes is an ubiquitous step in many problems across a variety of other fields, from Engineering and Computer Science (communication and information networks, database systems, information retrieval), to Biology, Sociology (biological and social networks) and Computational Linguistics [7–11]. In some sense, it can be regarded as one basic learning skill: Identifying two or more

*Electronic address: msantos@wodaklab.org; URL: http://wodaklab.org

distinct groups across an initial set of seemingly identical (or completely different) objects leads the way to investigate further details of each group. This view somehow encompasses the idea of emergent phenomena typical of complex systems. In computer science, more often than not clustering algorithms are found under the heading of (un)supervised *learning* algorithms, thereby explicitly highlighting its conceptual relevance. Furthermore, the concept itself of a partition may play a key role in analyzing different problems, even if their study do not involve an explicit clustering step. Particular examples of problems involving these concepts include drug discovery, in Chemistry; function determination of proteins based on their sequences, in Bioinformatics; optimal design of integrated circuits, in Engineering; the study of fragmentation processes in Nuclear and Condensed Matter Physics; first-order phase transitions in ferromagnetic materials; (bla,bla,bla). (NEEDS EXAMPLES/REFERENCES!!)

Each field has its own particular nuances leading to such a clustering problem. In Biology, it is a tacitly accepted principle that, in general, a protein sequence determines its function. This statement itself bears some additional difficulties as there is no definite consensus on what is meant by function. For instance, it could be understood at a molecular level, e.g. binding to DNA, at a metabolic level, affecting a chain of reactions, or at a phenotypic level. Despite such an issue, it is widely assumed that, considering also Anfinsen's dogma, all the information of the 3-dimensional structure and function of a protein is coded into its sequence (see [12] and references therein for a recent dispute of it). A fundamental problem in Bioinformatics is then to determine a protein's function from its sequence. The most common approach is finding an answer by comparison against a set of similar (homologous) proteins for which their function is known. A variant of this problem, however, is to determine subfamilies of slightly different specificities among a set of proteins with a common generic function. This is a clustering problem for which one can find almost as many different procedures as published articles about it, each considering the different additional information or algorithmic approaches the authors may consider fit.

Independently of the validity of the biological assumptions leading to such a clustering problem, the latter may still be of relevance, as one may ask about the amount of information on a protein's function one may extract from its sequence, even if function is not *uniquely* determined by the protein sequence *alone*. A clear formulation of the clustering problem and its solution may offer a solid ground upon which such question could be clarified. A similar conclusion may be drawn in other fields as well. Therefore, we will not further question the validity of the clustering problem, but assumed it as given.

Usually, the particular method used to solve a clustering problem depends on the details of the problem under study as well as on idiosyncratic preferences of the community or authors involved. In general, in the approaches hitherto found in the literature, the number of free parameters used is usually arbitrary and these lack a sound rationale for their inclusion; rather, they are justified *a posteriori* as those free parameters that allow one to obtain a pre-defined solution[7]. (MORE EXAMPLES/REFERENCES??). (*DOUBLECHECK: However, two general approaches, MCL [13] and Affinity Propagation [14], deserve a special mention. Affinity Propagation may have a sound formulation as a belief propagation minimizing the distance to a centroid, but its description lacks an easy intuitive view of the parameters used and their relevance in obtaining the optimal partition. Analogously, MCL, while similar in spirit - optimizing a markov flow on the network- does not really output an optimal clustering solution, as it lacks a scoring function to compare solutions against each other, but rather represents just a fast algorithm for sampling the partition space.*) Thus, to a large extent, the program set out in [15, 16] is seen as a reference yet: They all solve the *computational* problem of devising, by any reasonable means, an empirical method for classifying the given data[16].

As practical and legitimate as this approach is, we feel it is intellectually unsatisfactory. The field is lacking a comprehensive foundation explaining, among other things, (1) how to properly compare the solutions obtained taking into account the input information, (2) why and when it is reasonable to expect a boolean (crisp) partition instead of a continuous (fuzzy) one that accounts as well for overlap of communities; it lacks a general framework for (3) systematically deriving approximate (perturbative) solutions and for (4) allowing a systematic discussion of appropriate affinity measures $\boldsymbol{\omega}$ to use (see Fig.1).

There has been several attempts in the literature to derive a possible order-disorder like Hamiltonian for describing the clustering problem. However, we feel there is missing a comprehensive discussion on how to proceed. In particular, they all are based on ad-hoc assumptions on the elementary (spin) variables, parameters and the Hamiltonian to use [17, 18]. Neither has it been attempted to justify and derive from first principles a variational description of the clustering problem. From a practical point of view, such a variational approach seems a natural point to start from, specially among the physics community. Yet, one may question the generality of such descriptions as well as their range of applicability. Such discussions seem to be slowly taking off, e.g., on the modularity measure introduced by Newman and Girvan[19–22].

In this work we will concentrate on the problem of clustering a set of objects for which a pairwise (dis)similarity matrix $\boldsymbol{\omega}$ is available. We may also use the term *affinity* matrix in a more general context when we do not make precise if it is a similarity or disimilarity matrix. Thus, considering the possible meanings of the values of $\boldsymbol{\omega}$, we could make a somewhat naive distinction of three types of clustering problems: connectivity-based, bond-robustness and those of an attraction-repulsion system

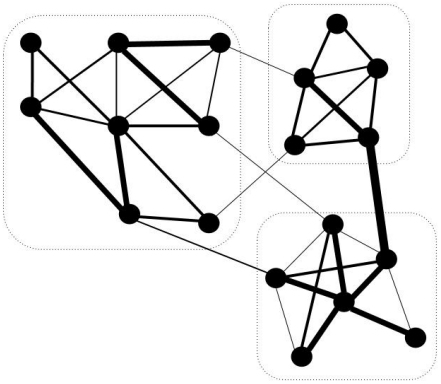


FIG. 1: The problem of clustering in a weighted network. The weights are assumed given by a pairwise matrix $\boldsymbol{\omega}$, called the *affinity* matrix, and may have varying meanings. Different edge weights are denoted by different edge widths. Edges are $1, 2, 3, 4, 8$ and 12 pixels in width. The vertices can be grouped into different clusters or communities depending on the criteria and information used. The spatial distribution of connectivities and vertices may easily lead the eye to distinguish the 3 clusters sketched. A topological criterion based solely on the connectivities of the vertices may be used in a computer program to identify the same three clusters by way of removing the edges with highest betweenness. However, a different algorithm that would consider the edge width as a kind of *bond strength* might well yield just 2 clusters: the 12px width edge, joining two of the former clusters, might be strong enough to consider both groups together as the optimal solution. In all cases, we can identify a property that is enhanced within each group, but damped out across groups: proximity (i.e., small distance) to a centroid, connectivity of vertices, and robustness of the structure of *bonds*. Conversely, the similarity of vertices is said to be higher within groups than across groups. Eventually, we may add a procedure for defining a probability distribution measuring the (continuous-valued) likelihood that each vertex belongs to a given group. This corresponds to a community structure as defined in the present work.

(see figures 1 and 2). Furthermore, in general we will focus on a *dense* graph of continuous-valued edges $\boldsymbol{\omega}$ for each pair of vertices (objects). Therefore, we will consider only in passing the case of sparse graphs where the only information available is the topological structure of connectivities between (a limited number of) vertices as in Fig.1.

B. Foundations for a General Theory of Clustering

Evidently, the possible foundations have to first be of a mathematical nature. The fact that the underlying mathematical structure is that of a complete lattice (i.e., a poset satisfying additional constraints) should remind us that any satisfactory description of partitions should not ignore it, but be compatible with such structure and try to convey as much information about it as possible[23–27], very much like in physics where the manifold where the observables live in impose some constraints on the mathematical laws describing the system [28].

We expect, however, that physics has also its say in this matter. The first link to physics comes from its relation to information theory[27, 29]. Zurek's algorithmic information distance defined in the space of all partitions entails the introduction of a probability distribution over the clusters of a partition [30–32]. We claim that the proper way to understand this probability distribution is as being conditioned on the *a priori* information considered when stating the given clustering problem. We are then lead to the question of how such a distance measure depends on the implicit *a priori* information. This is the first link we set out to unearth in the present work (see Fig.2).

Physics provides us with many examples of clustering processes in nature. Loosely speaking such cases could be described as processes of pattern formation in nonequilibrium systems. While the particular systems and their dynamics may be diverse [33, 34], they all show common trends. A particular interesting case is that of phase separation and the physics of the coexistence of multiple phases in a system [35]. While the bulk phase can be considered as homogeneous, particles of different phases diffuse into each other to some extent. At a mesoscopical level, the boundaries between phases are then rather continuous (or *fuzzy*, using a language from the field of clustering) instead of sharp. Such diffusion between phases gives rise to a superficial tension. Can we learn from these cases new approaches or guiding lines for use in the community-finding problem in complex networks? By analogy to the statistical mechanics of a binary system, we can define an average interface weight (energy), for scoring the optimality of a partition, that has been used in the literature before, albeit with questionable success. It might be worth, however, to further explore this analogy (see Fig. 2).

Besides these *physical* analogies, Statistical Mechanics may provide a more *technical* or *methodological* one. The density matrix approach seems to offer a very appealing new description of a partition. Following [36], "a pure ensemble by definition is a collection of physical systems such that every member is characterized by the same state $\mid i\rangle$. In contrast, in a mixed ensemble, a fraction of the members with relative population f_1 is characterized by $\mid 1\rangle$, some other fraction with relative population f_2, by $\mid 2\rangle$, and so on. Roughly speaking, a mixed ensemble can be viewed as a mixture of pure ensembles, just as the name suggests. The fractional populations are constrained to satisfy the normalization condition $\sum_i f_i = 1$". The density operator is then given

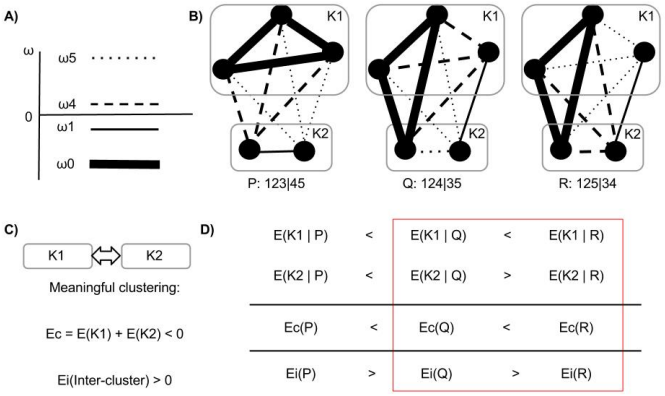


FIG. 2: **Physical analogy of clustering:** In nature, a phase separation of a mixture of particles splits the system into different phases or clusters depending on the interactions of the constituent particles. A) The interaction energy ω_{ij} between the particles (nodes) are represented by the edges and edge-widths. This example has two types of attractive interactions (negative energy), and two types of repulsive interactions (positive energy). B) Three different clusterings of 5 particles are given by partitions P, Q and R, all made of one big cluster $K1$ and a small one $K2$, each with varying content across the different partitions. C) The ideal meaningful clustering is assumed to have I) stable clusters (each intra-cluster compound energy is negative, $E(K1)$, $E(K2) < 0$), thus yielding a negative overall (intra-)cluster energy Ec, and II) a repulsive ($Ei > 0$) interaction among clusters. The latter may be called the criterion of discernibility of clusters. The total intra-cluster energy is then a measure of the stability (adequacy) of a clustering. D) Partition P is the optimal partition, while Q and R are seen as *approximations* to P. Cluster $K2$ is unstable in both partitions Q and R. Its energy is larger in partition Q than in R, but it is compensated by the energy of cluster $K1$ yielding a total cluster energy $E(Q)$ smaller than in R. Furthermore, the cluster interaction energy E_i (discernibility of clusters) is larger in Q than in R. Partition Q is therefore a better approximation to P than R. A sound distance measure that takes into account the edge-weights is expected to yield Q closer to R than P, i.e., $d(P,Q) < d(P,R)$. See text for details.

by

$$\rho \equiv \sum_i f_i \mid i\rangle\langle i \mid \tag{1}$$

where the diagonal matrix $\boldsymbol{f} = diag\,(f_1, f_2, \dots)$ describes the distribution of relative (cluster) populations. It seems evident that this description is tantamount to that of any clustering. Indeed, the latter always comes as a distribution of elements among a set of clusters, which here play the analog of states. More specifically, determining the population distribution described by the density matrix is akin to solving the number-theoretic problem of partitioning a non-negative integer into the sum of positive integers [37]. Below we will define a spinorial representation of a (*set*) partition that provides a geometric relationship between such number-theoretic partitioning of an integer and the equivalence relation characterizing the set partition. In seems then natural to consider the following question: Can we devise a description of partitions in terms of an equivalent density matrix and solve the clustering through an appropriate variational principle like in statistical mechanics? This is the second question that set forth the present work.

The late developments in the physics of information theory suggest that the clustering problem and the concept of partitions may have a more fundamental link to physics. When identifying clusters in the graph of Fig.1 our brain solves a problem (clustering a visual input) that is different from that a computer would perform when solving the same clustering question. In other words, the same partitioning question is answered using two different models of the data and two different physical substrates (the combination of the hardware, the program and the data can be thought of as a Turing machine processing a string of 0's and 1's from the input and the memory). In both cases, however, physical processes are involved that exchange energy and entropy with an environment[30, 31]. While arguably somewhat speculative, the question of how such processes may condition our choice of optimal clustering seems a sound and pertinent question.

Nevertheless, Physics provides us with a particularly powerful paradigm. That modern physics has unveiled the relevance of symmetries in the description of natural phenomena seems an understatement. Quantum Field Theory is essentially based on the identification of the Poincaré group as the one describing the symmetries of space-time [38] and the physics of Condensed Matter builds as well upon the concept of symmetry [39, 40]. Indeed, the early organization of the fundamental particles into families by Gell-Mann, Zweig and Ne'eman [41] could be viewed as an example where the concept of symmetry is used to guide the clustering of different objects into distinct groups [42]. Symmetries in Physics help *grouping* not only particles but also spectra. The *classification* of nuclear and atomic spectra is rooted in the symmetries of their corresponding Hamiltonians. The most familiar example of the latter is Mendeleev's periodic table of elements, which Quantum Physics explained by the *recurrent pattern* of the electronic configuration of atoms[43–45].

It seems worth the effort, therefore, to consider the relevance of symmetry in describing the community-finding problem in complex networks. Clearly, the implicit meaning of grouping a subset of elements together into a cluster is that we consider them all as *the same* with respect to a given property. In other words, in terms of our example of Homeodomains, wherever one such protein domain plays a given role, the same function could be carried out, *mutatis mutandi*, by any other sequence belonging to the same specificity cluster. The elementary symmetry we are describing here is that given by (a subgroup of) the Symmetric group $\mathbb{S}_N$ acting on the set of N elements (vertices of the network) to be

clustered. Any given partition is then isomorphic to a corresponding Young tabloid. The latter is defined as the equivalence class of Young tableaux invariant under permutations that shuffle only elements within the same row. In particular, any partition P can be adscribed a unique symmetry group $\mathbb{S}_P = \mathbb{S}_{P_1} \times \cdots \times \mathbb{S}_{P_K} \subset \mathbb{S}_N$ where each term $\mathbb{S}_{P_i}$ corresponds to the symmetric group of the corresponding cluster P_i of n_i elements.

We will show that there is, what currently seems, an alternative representation of partitions in terms of the symmetry group $SU(N)$ of unitary matrices describing rotations in N dimensions -or rather a subgroup of it. This is not simply the matrix representation of $\mathbb{S}_P$, but it unearths a (to our knowledge) novel relation between the number theoretic problem of partitioning an integer into the sum of integers and the problem of clustering a set of elements. Furthermore, the widespread criterion of ideal clustering on a graph is the vague idea that the similarity within clusters must be high, but that across clusters, must be low. In a few examples we will show that this criterion can be expressed as a symmetry condition, where the ideal case would correspond to a full spherical symmetry (in N dimensions), and where rotational symmetry around a given direction might be considered as cases showing only partial or approximate symmetries. In the latter case, symmetries might be used to identify an approximate solution with higher symmetry than the exact one, in an analogous way as it is done in Physics [46].

II. BAYESIAN APPROACH

A. The clustering problem

Consider the problem of clustering a set of homeodomain protein sequences. In general, clustering algorithms cannot cope with all possible homeodomain sequences N' and we are given a subset $N < N'$ of those sequences with the hope that for large enough N the results will be statistically significant or robust enough so that *small* changes should be expected if we would add the rest of the sequences $N' - N$. Indeed, we would like the identified homeodomain clusters to represent actual, distinct homeodomain properties, e.g. distinct specificities. The conviction that these distinct specificity classes manifest in nature as distinct phenotypic processes lead us to confer those clusters a real content beyond the particularities of the clustering algorithm and data set used. In other words, we can expect some clusters to change when varying N, but we assume we can device a clustering process that, for large enough N, identifies an emergent *pattern of clusters* that would not change when adding further sequences. (This statement needs further precision as to under which conditions can we expect it to hold if we add an arbitrary number of arbitrarily different sequences. We will try to clarify this point along the rest of this work). Each cluster, thus, represents a particular (discrete) value of a physical observable (or, rather, biological in this example), namely, the specifity or functional class of a given homeodomain.

This conclusion does not change, however, if we would have access to an algorithm capable of clustering all presently known homeodomain sequences. This is so for two reasons: First, because there is implicit a *bisection* of the space of sequences into homeodomain and non-homeodomain sequences. This is a preliminary clustering step that is assumed among the *a priori* information $\mathcal{H}$ available and that is subjected to the same nuances inherent to any clustering solution. Second, because no clustering process is guaranteed to give the *true* solution, but rather a solution in terms of communities or *fuzzy* partition where the *boolean* counterpart (each element belongs to one and only one cluster) always comes after a subjective definition of where to *cut* the network (see Fig.1).

This last statement has a discrete counterpart as well. Let us assume that all clusterings have to be of a discrete (boolean, as opposed to fuzzy) nature, where each element belongs to one and only one cluster. We would rephrase the final statement of the last paragraph by claiming that we can expect at most to obtain a family of partitions all equally valid as solution to our original clustering problem. The fuzziness could then be introduced by way of assigning probabilities to each cluster.

Incidentally, this last picture entails a limit on the resolution of the answer we can obtain (depending on the *a priori* information $\boldsymbol{\omega}$), which could be viewed as a fluctuating nature of the solution of the clustering problem. Talking about fuzzy partitions or communities would then be just a reflection of the degeneracy of this solution. This raises the following question: Are there clustering problems where the genuine solution is intrinsically a fuzzy partition? or are all fuzzy partitions just the result of an intrinsic multiplicity of solutions of the clustering problem -like, e.g., a system with a complex energy landscape?

This question is reminiscent of that raised by the advent of Quantum Mechanics. Is the probabilistic description an artifact of our limitations in devising an algorithm that gives a unique answer? or is it rather intrinsic to the clustering problem, either from an epistemological or, most likely, an algorithmic point of view? Does algorithmic complexity theory have anything to say in this respect? Recent works in the literature have argued that the description in terms of probability amplitudes we find in Quantum Theory is but a necessary consequence of the symmetries of some basic logical assumptions in the structure of the theory[24, 25]. Does the lattice structure of the space of partitions bears such intrinsic probabilistic description? Could we carry over their argumentation to our case?

The link to Quantum Theory may seem far-fetched at first. However, given our goal of finding an as general as possible foundation of the clustering problem the results by Goyal et al. [24, 25] provide a reasonable ground

for discussing the nature of the probability distribution that arises in the community-finding problem and the determination of a sound metric for it. The motivation for these questions lies on a seemingly unavoidable constraint imposed by a desirable theoretical condition of normalization. In this work we will describe what is a fundamental relation between the matrix representation of an equivalence relation and the number-theoretic problem of partitioning an integer into a sum of integers. If we introduce a continous-valued membership function describing the communities of a fuzzy partition, and impose two conditions, namely, (1) that the fundamental relation holds as well for fuzzy partitions and (2) that the equivalence matrix describes an integer number N of elements, then we seem to come to the conclusion that it is the *square of the membership functions* that describe the density of elements in each community, not those functions themselves. This is reminiscent of the algebraic structure of the description of states in Quantum Mechanics, where the square of the probability amplitude respresents a probability for finding a particle in that state.

B. On the nature of *states*, *particles* and *temperature*

Physics has been quite successful in deriving a probabilistic, but intuitive description of systems composed of many degress of freedom. Statistical mechanics deals with (macroscopic) systems (in equilibrium) which have many possible realizations (microscopic states) available, each satisfying the same (macroscopic) conditions of the system. Under the external constraints imposed by a heat bath, a particle will hop across different micro-states compatible with the constraints. The hopping of one particle is then tantamount to the hopping of the system from one point in phase space to another (possibly equivalent from a macroscopic viewpoint).

As discussed in subsection II A, among the *a priori* (but, usually, implicit) information the clustering problem relies on is the idea of an objective existence of the groupings determined by a clustering procedure. These clusters, or moreover, the property they *identify* is reminiscent of the macroscopic constraints mentioned above (see Fig.2). It seems, therefore, not far stretched to try stablishing a (formal) link between Statistical Mechanics and our clustering problem [17, 47, 48].

As we set out to borrow ideas, concepts and methods from Physics in general, and in particular, from Statistical Mechanics, it seems inevitable that we face some obvious questions related to the seemingly, drastic difference in nature of the problems and systems studied in Physics and those in the discipline of clustering. The three main issues are understanding the equivalence of the physical concepts of states, particles and temperature. This would be essential in understanding the applicability of formalisms of Physics in the clustering problem. However, by no means should this analogy be considered as a claim for ascribing any ontological meaning to these concepts in the context of the clustering problem. Rather we see it as a possible, convenient and fruitful analogy that may allow a more comprehensive approach and understanding of the clustering problem. We will start with addressing the first two concepts.

Let us consider a few examples in other to motivate the discussion. In set theory, when we group together several elements forming a class, we mean that each of them satisfy a common set of conditions, namely, the predicates or properties that characterize the given set. For example, the set of natural numbers $\mathbb{N}$ satisfy Peano's axioms; the set of all even integers can be characterized by $(2) \equiv \{2k; k \in \mathbb{Z}\}$; a 3-dimensional unit sphere is the set of points satisfying $x^2 + y^2 + z^2 = 1$; the set of all mouse homeodomain proteins have a common function (chiefly, the regulation of patterns, or morphogenesis) in the early stages of embryo and cell development; or the set of all atoms of a given tile on the floor. From the perspective of the common properties, all elements in each of these sets can be considered as equal. Alternatively, each element can be seen as a different instance of the same *concept* of element: a point on a 3-dimensional unit sphere, the natural number 17, the mouse homeodomain $Hoxa12$, or a tile's atom, each can be seen as a particular sampling of the whole set of corresponding elements. According to current ideas in Evolutionary Biology, this is what happens in nature where, during evolution, gene duplication and speciation events allow nature to sample different realizations of a gene family (e.g., homeodomains) -although not all genes are equally abundant across different species, or alternatively, not all are equally fit for a given species.

The question we are raising is then the following: Are elements of a set that are to be classified akin to different *states* of a system? Or rather should those elements be seen as *particles* of a model of N-*body* system, where a partition would be a state of such a system, thus akin to an instance of a statistical ensemble of particles? The precise clustering would then correspond to a specific configuration of the system and the space of all configurations would be $\mathbf{\Pi}(S)$, the space of all partitions of a set S, thus playing the role of the *phase space* of S. Each element would then be a particle and each cluster would correspond to the value of an observable like the energy or spin.

Alternatively, we might consider the elements of S as its possible states, that is, the elements themselves represent possible configurations or *microstates* of S and any given clustering would be akin to hyperplanes in phase space satisfying specific energy or symmetry constraints.

In both cases, the concept of cluster could share a common meaning. Physics defines a particle species as the set of states of an irreducible representation of the Poincaré Group [41]. Can a cluster be considered as a *particle* species in a similar way, i.e., as an irreducible representation of a suitable symmetry group? What would this symmetry group be? In this work we claim to shed some

light on this last question.

Turning back our attention to the first possibility, if we consider $\mathbf{\Pi}(S)$ as the space of (clustering) *states* or phase space of S, what would then the ground state be? The motivation for the last question comes from realizing the fact that any partition can be derived by a sequence of clusters-merging steps starting from the partition where all elements constitute their own cluster (*singleton* cluster). Merging a singleton with another cluster is then like leaving an empty cluster, or *creating a hole*, and promoting the system to an *excited* state. Thus the partition $P = 12|3 = 12|\cdot|3$ could be viewed as the result of merging singleton 2 with 1, starting from partition $\bar{0} = 1|2|3$, and leaving one empty cluster (*hole*) represented by the dot. Can we consider the partition $\bar{0}$, that of all singletons, as the ground state in the clustering of S? That is, can we set up a suitable Fock space for describing such a ground state and its excitations? We will see below, when defining an (Shannon-like) entropy for a partition, that $\bar{0}$ is the partition with maximum entropy -if we ignore the information provided by $\boldsymbol{\omega}$. Of course, as $\mathbf{\Pi}(S)$ is finite, we could always redefine this entropy such that the partition $\bar{1} = S$, the one clustering all elements together, has maximum entropy instead. We could then discuss the suitability of considering $\bar{1}$ the ground state of S[26, 27].

Incidentally, such a view raises an interesting question: each step increasing the number of clusters, is it also increasing the order of the system -i.e., decreasing the symmetry of the system? And viceversa. This is certainly a seemingly intuitive and appealing idea if we consider the analogy of a cluster as a hiperplane in the phase space of S given by all states (elements) in that cluster. This would be akin to the case of pattern formation or phase separation processes in physics. However, this idea seems at odds with the concept of an entropy having its maximum at $\bar{0}$. The contradiction seems to arise from two assumptions. The first is the meaning of entropy we are using, namely the amount of information needed in orther to classify any given element, i.e., to identify the cluster it belongs to. The more clusters a partition has the more yes/no questions we would need to answer in orther to correctly group that element. Thus, this entropy does not emcompass the idea of ordering, but only that of the intrinsic (algorithmic) complexity of the partition. The second is the implicit assumption of the elements of S as different *states* the system may be found in. Once two distinctive clusters are defined, the sampling (hopping) would take place either in one or the other cluster. This certainly would decrease the usual idea of entropy, therefore increasing the order of the system. Can we give such idea of states a solid ground? can we make it compatible with that of particles? In short, are the elements of S *states* or *particles*? This means, should we consider the elements of S states the system samples, with the clusters limiting ergodicity to a subset depending on some conditions to be clarified? or do the elements of S together *with* the cluster divisions define a specific state of S -or in short, is $\mathbf{\Pi}$ the *phase-space* of S?

If we want to make a link with Statistical Mechanics, we will then need to define an additional entropy source (ensemble entropy) describing the ordering state. This new term should be added to the algorithmic entropy, describing the intrinsic randomness of the available information given by the partition itself. Both together would thus yield the total entropy suitable for our goal [30, 31].

Obviously, there is a hierarchy of solutions of different granularity when clustering a group of items. Each of the above listed sets, in turn, can be considered as part of a larger class containing similar groups, the latter constituting a partition of all elements in the class: all tiles of the floor group the floor's surface atoms into distinct clusters (the tiles), thus allowing us to make a distinction between atoms on one tile and those on another. Similarly, all homeodomains belong to the broader set of transcription factors, proteins which regulate gene expression. At the same time, however, some mouse homeodomains may regulate the brain development while other are responsible for the anterior/posterior pattern formation, and each of these groups showing a preferential binding to different DNA motifs -different binding specificity. Thus, we can also consider a finer subdivision of an initially equivalent set of elements.

{REVISE/REMOVE} Under this perspective, how does the sampling picture change? In addition to gene duplication and speciation events nature encompasses gene extintion events, e.g. with the extintion of a whole species. Thus, the first two events allow nature to sample the available gene space by creating new sequences (adding new elements to S) while the last one simply eliminates sequences (reducing the number of elements). The number of elements N is thus fluctuating. Although at any point in evolution we may consider it a specific number, the time scale of gene mutations, even if these are neutral ones, is small enough for the observations in Section II A to be of relevance.

In all cases, all elements are *de facto* seen as different *states* of one *particle* or *system* -the whole set. These states may all be equivalent, e.g. when considering the trivial partition given by the whole set itself (i.e., no partition), or not, like in the case of a non-trivial partition of the original set, as e.g. when splitting it into two disjoint classes.

energy states if the temperature is high enough, but will sample only equivalent micro-states of equal energy otherwise.

The third fundamental concept we would like to borrow from Physics is that of temperature. How does the concept of temperature in Physics fits into this picture? Condensed Matter Physics gives us plenty of examples where a gradual variation of temperature unveils a hierarchy of new symmetry states of a system. As an example, we may consider the following sequence of phases: fluid state, a smectic-A liquid crystal (Sm-A), a Sm-C, a Sm-F, a Sm-G, and finally a crystalline solid. Such a

sequence of thermotropic changes, gradually decreasing the symmetry of a system, is found, e.g., in a material composed of the aromatic chain 2-(4-n-pentylphenyl)-5-(4-n-pentyloxyphenyl)-pyrimidine [39]. Can we consider the affinity matrix $\boldsymbol{\omega}$ as akin to *interaction energy*, i.e., a *Hamiltonian*? If so, what plays the role of intrinsic noise, i.e., what is the source of *temperature*?

Hitherto in the literature, the parameter controlling the granularity of a partition has been implicitly considered the equivalent of the thermodynamic temperature. This, however, is somewhat misleading, as it always takes a value depending on the rather subjective criteria of the author. Furthermore, given the current lack of a unifying formalism for clustering, it seems unlikely to find a way for giving a meaning and comparing this granularity parameter between different systems using the same clustering algorithm, and even less when it comes to comparing results of different, independent studies.

However, previous work on the effects of adding additional (protein) sequences to a previously classified set (partially unpublished) may shed some light on this question. Depending on the relative number of new sequences added and on their sequence similarity (against each other and against the original set), results show the breaking up of the original clustering in a way that, for this example at least, suggests a similar interplay between sequence identity ($\boldsymbol{\omega}$, understood here as *interaction energy*) and number of additional sequences (*noise*?). Could such a test allow for the definition of an *intrinsic temperature* of a set of sequences? Say, e.g., by way of quantifying the easiness with which the original system gradually breaks up. It seems at least worth to further explore these ideas.

Last remark: if we consider the clustering of a set by a suitable turing machine, can we device a (gedanken) experiment showing the relevance of the number of sequences and relative identities as the source of an actual, in the sense of physical, temperature? What is the nature of the radiation emited? what is the mechanism of radiation?

C. On the nature of the affinity measure $\boldsymbol{\omega}$

In [30, 31], and more in detail in [49], a discussion is presented on suitable measures of similarity relevant to pattern recognition based on algorithmic-theoretic information distances. Three main requirements are identified [49] for any suitable similarity measure: 1) Similarities between objects can be represented by effectively computable functions or at least (upper/lower) semicomputable functions of binary strings. In layman's words, there has to be a deterministic and finite computer program capable of calculating the similarity value between any pair of objects; alternatively, the similiraty function is the limit of a computable sequence of upper/lower bounds. 2) For each object x and (di-)similarity value d there are finitely many objects y at a (di-)similarity values d. This requirement is conveyed by the following *normalization* condition

$$\sum_{y:y\neq x} 2^{-|\boldsymbol{\omega}_{x,y}|} \leq 1. \tag{2}$$

The rationale for this inequality is grounded on analogy with the Hamming distance between two binary strings: There are exactly 2^d strings within a Hamming distance d from any given string. Lastly, there is the requirement that 3) the similarity measure be a metric [49]. In algorithmic information theory, the meaning of the distance metric $E(x,y)$ between two binary strings x, y is given by the length in bits of the shortest program p that can output y given x and x given y. (additional conditions are assumed as that the program be executed by an universal Turing machine and that a self-delimiting code be used for transcribing any string to binary format). The shortest programs each computing the conversition in only one way are not longer than p, with lengths denoted by $K(y|x))$ and $K(x|y))$, respectively. The *admissible* distances are then defined as upper-semicomputable, normalized metrics. Two possible alternatives are $E_1(x,y) = \max\{K(y|x), K(x|y)\}$ and $E_3(x,y) = K(y|x) + K(x|y)$, the latter being Zurek's algorithmic information distance and is essentially the one used in subsection II D. We shall use in this section the same notation as in the above mentioned references. The other distance function $E_2(x,y) = K_R(y|x) = K_R(y|x)$ refers to completely reversible transformations, which we will not consider here. In [49] it is proved that E_1 is *universal*, in the sense that it is a lower bound to any other possible admissible distance. Moreover, it is $E_1(x,y) \leq E_3(x,y) \leq 2E_1(x,y)$ [49]. In addition, two definitions of thermodynamic entropy cost of transforming x into y by the most efficient process have been proposed [30, 49] as $\delta S(x \to y) = K(x) - K(y)$ and $\delta S(x \to y) = K(x|y) - K(y|x)$.

How does our prior information $\mathcal{H}$, and in particular $\boldsymbol{\omega}$, fits into this picture? Is $\boldsymbol{\omega}_{\alpha\,\beta}$ a partial *insight into* or estimate of $E(\alpha,\beta)$? or maybe an estimate of $\delta S(\alpha \to \beta)$? If we are looking to draw an analogy with Statistical Mechanics, is it sound to expect $-\log P(\alpha \sim \beta|\mathcal{H}) \sim \boldsymbol{\omega}_{\alpha\beta}$? How would it give rise to such a probability distribution? What happens if we are given E_1 or E_3 as prior information $\boldsymbol{\omega}$? How does the clustering change, or more interestingly, the comparison of partitions? How does the ensemble picture given by those probabilities emerge? For instance, can any such distances $E(\alpha,\beta)$ define a cluster radius, or give rise to a *density* of elements in each cluster that links to $P(\alpha \sim \beta|\mathcal{H})$? Can Algorithmic information theory explain why $E_3(P,Q)$ defines a geodesic containing $P \wedge Q$?

Bennet et al.[49] proved the following exponential relation. Be $B_i(x,d) \equiv \{y;\, E_i(x,y) \leq d\}$, the ball centered in string x of radius d as given by the distance E_i $(i = 1, 3)$, and $|B_i|$ the number of elements (strings) in such a ball. Then it holds $\log |B_i(x,d)| = d - K(d|x)$, with logarthmic corrections for $i = 3$. In general, we expect the clusters

to consist of elements that do not differ *too much*. Thus, if the prior information comes in the form of a disimilarity matrix given by algorithmic distances $E(\alpha,\beta)$, the result of Bennet et al. suggests that, on average, the likelyhood of finding two randomly chosen elements α and β within the same cluster depends exponentially in the prior information $\boldsymbol{\omega}_{\alpha\beta} = E(\alpha,\beta)$.

Temptative more precise arguments: Be a partition $P = \{P_1, \cdots, P_{\mathcal{K}}\}$, with a radius for each cluster defined by $2\,R_i \equiv \max\{E(x,y); x,y \in P_i\}$. According to the above discussion, we can always find a $c_i \in P_i$ for which $\log|P_i| \simeq 2R_i - K(R_i|c_i)$. The probability of finding two randomly chosen elements x and y within the same cluster is

$$\begin{aligned}\mathcal{P}(x\sim y|P,\,\mathcal{H}) &= \sum_{i=1}^{i=\mathcal{K}} \mathcal{P}(x\in P_i|P,\,\mathcal{H})\,\mathcal{P}(y\in P_i|x\in P_i,\,P\,\mathcal{H}) \\ &= \sum_{i=1}^{i=\mathcal{K}} \frac{n_i}{N}\,\mathcal{P}(E(x,y)+E(x,c_i)<R_i|P,\mathcal{H}) \\ &\sim \sum_{i=1}^{i=\mathcal{K}} --\end{aligned}$$

where x, y are two arbitrary elements; alternatively

$$\begin{aligned}\mathcal{P}(x\sim y|P,\,\mathcal{H}) &= \mathcal{P}(x|P,\,\mathcal{H})\,\mathcal{P}(y\in P_{P(x)}|x,P,\,\mathcal{H}) \\ &\sim N^{-2}\,e^{(2R_x-K(R_x|c_x))} \\ &= N^{-2}\,e^{(2R_y-K(R_y|c_y))} \\ &= N^{-2}\,\frac{1}{2}\,\{e^{(2R_x-K(R_x|c_x))} + e^{(2R_y-K(R_y|c_y))}\} \\ &= N^{-2}\frac{1}{2}\,e^{(2R_x-K(R_x|c_x))}\cdot \\ &\quad\cdot\ \{1+e^{[2(R_y-R_x-(K(R_y|c_y)-K(R_x|c_x)))]}\}\end{aligned}$$

where $R_z \equiv R_{P(z)}$; or alternatively

$$\begin{aligned}\mathcal{P}(x\sim y|\mathcal{H}) &= 1-\mathcal{P}(x\nsim y|\mathcal{H}) \\ &\lesssim 1-\mathcal{P}(E(x,y)>R_x|\mathcal{H}) \\ &\lesssim 1-\mathcal{P}(E(x,y)>R_y|\mathcal{H}) \\ &\sim 1-\frac{1}{N}\,\frac{(1-n_{P(x)})}{N}\end{aligned}$$

But these expressions do not seem to relate to $E(x,y)$ like, f.i.

$$\begin{aligned}&n_{P(x)}n_{P(y)} \sim \\ &\sim N^{-2}\,e^{2(R_x+R_y-\frac{K(R_x|c_x)+K(R_y|c_y)}{2})}\end{aligned}$$

which for overlapping clusters $E(x,y) \leq R_x + R_y$, but for non-overlapping the inequality is inverted...Not sure if I can get somewhere...

What is the most general transformation $\boldsymbol{\omega}_{\alpha\beta} = F(E(\alpha,\beta))$ that preserves such an exponential relation for $\boldsymbol{\omega}_{\alpha\beta}$? For instance, $F(x) = x + a$, for any constant $a \geq 0$, is such a function. From geometric formulations of Thermodynamics [50, 51], we expect the probability for a system to randomly jump from one state to another to decay exponentially with the entropy difference between these two states. Is such a relation equivalent to the one just discussed above? Can a general affinity matrix $\boldsymbol{\omega}$ given by a transformation F be interpreted as an (transition) energy or hamiltonian for the purpose of clustering? May legendre transformations be relevant here?

D. Comparing partitions

In the clustering problem, implicitly, we consider that any partition is always defined relative to an underlying property or hypothesis $\mathcal{H}$, much like the concept of an a priori knowledge in Bayesian approach to Probability theory. Consider the problem of clustering a set of protein domains into distinct specificity classes, e.g., the so called homeodomains, being identified as such chiefly because all bind to the DNA in order to control the early stages of embryo development. In this respect, they all are the same, namely homeodomains. Yet, some of them are responsible for the development of different limbs, others control the skeletal system development, other yet that of the central nervous system, or the cerebellum, etc. Specificity is then a property than can take different values, much like an observable in physics. Given this prior information, which includes the existence of different specificities, we can ask about the probability of any two homeodomains A, B sharing the same specificity $A \sim B$, which we will denote here as $P(A\sim B|\mathcal{H})$. This probability will definitely depend on their sequence similarity, but eventually it may also depend on other, as of yet unknown properties that may or not be derived from the sequence similarity of both homeodomains. For instance, the chemical properties along the protein backbone can be derived from the individual amino acids, but the protein folding problem is far from being solved yet. For the sake of simplicity, let's interpret this similarity ω between any two elements A and B such that

$$-\log P(A\sim B|\mathcal{H}) = \boldsymbol{\omega}_{AB} + \mathcal{O}(\boldsymbol{\omega}_{AB}), \qquad (3)$$

where $\boldsymbol{\omega}_{AB}$ is a pairwise *affinity* or similarity value between elements A and B and is assumed to be given among the prior information $\mathcal{H}$.

The original problem motivating the present work was to define a distance between partitions that takes into account the affinity values $\boldsymbol{\omega}$, i.e., it's not only *combinatorial/entropic* as is Zurek's algorithmic information distance [30] (later rediscovered by Meilă in the context of partitions as Variation of Information or VI distance [32]). The entropy defined in that work encompasses only the algoritmic randomness implicit in the very same description of a partition for the reasons discussed above. An example may clarify this point.

Let us consider the following three partitions of $N = 5$ elements into 2 clusters (see Fig.2), $P = 123|45$, $Q =$

124|35 and $R = 125|34$, with a similarity matrix such that

$$0 \geq \omega_0 \equiv \boldsymbol{\omega}_{12} = \boldsymbol{\omega}_{13} = \boldsymbol{\omega}_{23} \neq \boldsymbol{\omega}_{45} \equiv \omega_1 \leq 0 \quad (4a)$$
$$0 \leq \boldsymbol{\omega}_{4i} \ll \boldsymbol{\omega}_{5i} \; i = 1, 2, 3 \quad (4b)$$

Intuitively, it seems appealing to expect that a sound distance measure among partitions would determine Q to be closer to P than R is, as the latter condition states that element 4 is more similar to any of elements $1, 2, 3$ than 5 is.

The VI distance between two partitions P, Q is defined as [30, 32]

$$d(P, Q) \equiv 2\, H(P \wedge Q) \; - \; H(P) \; - \; H(Q) \quad (5)$$

where $P \wedge Q$ is the intersection or meet of both partitions (see section III) and $H(P)$ is the Shannon entropy of a partition defined as

$$H(P) \; \equiv \; -\sum_k p_k \, \log \, p_k \; = \; -\frac{3}{5} \, \log \frac{3}{5} \; - \; \frac{2}{5} \, \log \frac{2}{5}$$
$$= \; \log 5 \; - \; \frac{3}{5} \, \log 3 \; - \; \frac{2}{5} \, \log 2$$

with $p_k \equiv \frac{n_k}{N}$ and n_k is the number of elements in cluster k. As only cluster-size distribution matters, in our case it is $H(P) = H(Q) = H(R)$. Also, we have

$$P \wedge Q \; = \; P \wedge R \; = \; 12|3|4|5 \quad (6)$$

Thus, according to the VI distance, $d(P, Q) = d(P, R) = \frac{6}{5} \, log \, 3$ and we would be inclined to consider R as good an approximation to P as Q is, something which feels at odds with our intuition given the prior information provided by the affinity matrix $\boldsymbol{\omega}$.

From the point of view of Algorithmic Information Theory[30], the distance $d(P, Q)$ yields the total amount of information needed to carry out the replacement of P by Q or vice versa by means of a reversible process. What we claim here is that, with the given information $\mathcal{H}$ (which includes $\boldsymbol{\omega}$), we cannot justify putting on an equal footing both partitions Q and R *relative* to P. On the contrary, *we would need an additional information to justify such claim* explaining why we should consider Q and R equally close to P despite being $\boldsymbol{\omega}_{4i} \ll \boldsymbol{\omega}_{5i}$. That is the meaning we expect a sound distance measure to convey.

E. The information theoretic approach

In these notes we examine the possibility of providing a unique theoretical framework, based on ideas of Statistical Mechanics, that could provide us with both a well defined algorithm as well as a sound justification of the free parameters involved. Furthermore, by construction, the latter would be the minimal number of free parameters needed in order to determine the optimal clustering.

In general, the clustering problem can be defined in terms of a set of elements $S = \{s_\alpha\}$, $\alpha \in [1, \cdots, N]$ and a pairwise affinity measure (e.g. in terms of pairwise dis/similarities) $\boldsymbol{\omega}_{\alpha\,\beta}$ defined on S. The goal is then to identify the set of equivalence classes of elements that is most *consistent* (in a way to be precised) with the given measure $\boldsymbol{\omega}_{\alpha\,\beta}$.

Any partition P of S can be described by an $N \times N$ matrix $\mathbf{A}$, called the adjacency matrix, which, for each pair of elements α, β, specifies whether they belong ($\mathbf{A}^{\alpha\beta} = 1$) or not ($\mathbf{A}^{\alpha\beta} = 0$) to same cluster. We are not aware, however, of any previous work in the literature implementing a full-fledged matrix, or vectorial, representation of a partition, that is, one mimicking the algebraic operations of union and intersection of partitions. Given such a representation, to any given partition P we could, in principle, ascribe a vector as

$$\mid P\rangle \; = \; \sum_{\alpha i} \mathbf{a}^{\alpha}{}_{i} \; \mid \alpha\rangle \mid i\rangle, \quad (7)$$

where $\mid k\rangle$ represents the k-th cluster, and $\mid \alpha\rangle$ represents a node (*state*) of the set of elements (*system*) we want to describe.

Up to this point, (7) is just an ad-hoc representation. In order to have a complete sense we would have to prove that the set of all vectors $\mid P\rangle$ really span a vectorial space, i.e., we would need to define what is meant by the *sum* of two partitions $\mid P\rangle + \mid Q\rangle$, as well as the *multiplication* by a scalar $\lambda \mid P\rangle$. While, mathematically, both operations are well defined their interpretation in the present context in not clear. We will postpone this discussion for later however. The aim, here, is rather to explore the possibility of an alternative description of a partition on a graph in terms of the usual formalism of Statistical Mechanics.

The structure of this draft is as follows: in section III we present some basic definitions related to the algebra of partitions, discuss their representation problem and introduce some useful concepts for describing partitions defined on a graph.

III. BASIC DEFINITIONS

A. Algebra of Partitions

A (*boolean* or *crisp*) *partition* P of a set of N objects $S = \{s_\alpha\}$ (data set) is defined as a collection $\{P_k\}$ of non-empty, mutually disjoint subsets of S (*clusters*) that cover the whole data set S. Here on we will use Latin characters when referring to cluster indexes and Greek letters for those of elements of S. Formally,

$$P \; = \; \{P_k\} \; ; \; \forall k \neq l \, , P_k \cap P_l \; = \; \emptyset \quad \text{and} \bigcup_k P_k \; = \; S.$$

We denote by $\|P\|$ the number of clusters in P and by $\|P_k\|$, the number of elements in cluster P_k. Specifying

a partition P on S is tantamount to defining an *equivalence relation* on S, which consists on the set of pairs of elements of S, $\mathbf{A} \equiv \{(s_\alpha, s_\beta)\} \subset S \times S$, such that s_α and s_β are considered as belonging to the same cluster, and such that $\mathbf{A}$ satisfies three essential conditions, namely to be *reflexive*, *symmetric* and *transitive*, i.e., for any three elements $s_\alpha, s_\beta, s_\gamma$ it is

$$\forall s_\alpha \;\Rightarrow\; (s_\alpha, s_\alpha) \in \mathbf{A} \tag{8a}$$
$$(s_\alpha, s_\beta) \in \mathbf{A} \;\Rightarrow\; (s_\alpha, s_\beta) \in \mathbf{A} \tag{8b}$$
$$\left.\begin{array}{l}(s_\alpha, s_\gamma) \in \mathbf{A}\\ (s_\gamma, s_\beta) \in \mathbf{A}\end{array}\right\} \;\Rightarrow\; (s_\alpha, s_\beta) \in \mathbf{A} \tag{8c}$$

$\mathbf{A}$ represents the property or *feature* that any two elements share in common if they are to be considered as belonging to the same cluster.

One defines that a partition Q is a *refinement* of partition P, $Q \trianglelefteq P$, or, equivalently, that P *covers* Q, when all clusters of Q are subsets of clusters in P, that is,

$$Q \trianglelefteq P \Leftrightarrow \{\forall\, Q_l \in Q \Rightarrow \exists\, k\,;\, Q_l \subseteq P_k\} \tag{9}$$

where $\{P_k\}$ and $\{Q_l\}$ are the collection of clusters defining partitions P and Q respectively. The symbol $\triangleleft$ will denote the strict refinement, i.e., excluding the equality. We will denote the set of all partition on a set S by $\Pi(S)$ or simply Π if there is no ambiguity. The relation $\trianglelefteq$ defines a partial order in $\Pi(S)$. Furthermore, the structure $(\Pi(S), \trianglelefteq)$ is a (complete) lattice. An immediate consequence of this is that there exists two elements in Π, $\bar{0}$ and $\bar{1}$, such that $\forall P \in \Pi$

$$\bar{0} \trianglelefteq P$$
$$P \trianglelefteq \bar{1}.$$

Thus $\bar{0}$ corresponds to the partition where all elements are *singletons*, i.e., constitute their own cluster, and $\bar{1}$, to that where all elements of S are clustered together, i.e., $\bar{1} = S$. Therefore, it is

$$\|\bar{0}\| = N \;\;;\;\; \|\bar{0}_k\| = 1$$
$$\|\bar{1}\| = 1 \;\;;\;\; \|\bar{1}_1\| = N.$$

Furthermore, the partial order induces two operations on $\Pi(S)$, the intersection $\wedge$ and the union $\vee$, which together characterize completely the algebraic properties of the lattice. For further details on the algebraic properties of lattices we refer the reader to any standard textbook on algebraic lattice theory [52, 53].

The intersection of two partitions P and Q is defined as the set of all non-empty intersections among their clusters, or more formally,

$$P \wedge Q \;=\; \{P_k \cap Q_l\,;\, P_k \cap Q_l \neq \emptyset\} \tag{10}$$

From 10, we have $\forall P, Q \in \Pi$,

$$P \trianglelefteq Q \Leftrightarrow P \wedge Q = P \tag{11}$$

and therefore $\forall P \in \Pi$

$$\bar{0} \wedge P \;=\; \bar{0}$$
$$\bar{1} \wedge P \;=\; P$$

which justify the notation used for these two partitions.

The union of two partitions P and Q, $P \vee Q$, is defined as the smallest partition that cover both P and Q.

For the sake of simplicity, hereafter, and without any lost of generality, we will identify the set of elements S with the set of integers $\{1, \cdots, N\}$. A partition P on S defines uniquely the *canonical* mapping

$$\begin{aligned} P :\, S &\to \{1, \cdots, \mathcal{K}\} \subset \mathbb{N} \\ \alpha &\mapsto i = P(\alpha) \end{aligned} \tag{12}$$

where $\mathcal{K} = \|P\|$. That is, the mapping P assigns to each element α the equivalence class $P(\alpha)$ where it was grouped according to the partition P. Similarly, a cluster P_k of a partition P on S defines uniquely the mapping

$$\begin{aligned} P_k :\, S &\to \{0, 1\} \subset \mathbb{N} \\ \alpha &\mapsto P_k(\alpha) = \delta_{k\,P(\alpha)}, \end{aligned} \tag{13a}$$

where $P_k(\alpha)$ is the *membership function* of cluster P_k, and we can write the latter as the set

$$P_k \;=\; \{s_\alpha \in S \quad ; \quad P_k(\alpha) = 1\}. \tag{13b}$$

The concept of membership function allows for an alternative definition of partitions that is easier to generalize beyond the boolean case. In the general case, which we will indistinctly call a *continuous* or *fuzzy* partition, the membership function takes continuous values in the interval $[0, 1]$,

$$P_k(\alpha) \;=\; \boldsymbol{\delta}_{k\,\alpha} \;\in\; [0, 1] \subset \mathbb{R}. \tag{14a}$$

The algebra of partitions defined above still holds in this case provided that we define appropriately an inclusion and intersection of fuzzy sets. We will adopt the following definitions

$$P_k \subseteq Q_l \;\Leftrightarrow\; \forall\, s_\alpha \in S\,;\, \boldsymbol{\delta}^P_{k\,\alpha} \leq \boldsymbol{\delta}^Q_{l\alpha} \tag{14b}$$
$$\boldsymbol{\delta}^{P_k \cap Q_l}_{i\,\alpha} \;\equiv\; \boldsymbol{\delta}^P_{k\,\alpha}\, \boldsymbol{\delta}^Q_{l\,\alpha} \tag{14c}$$

where the latter defines the membership function corresponding to the intersection of two clusters P_k and Q_l. In addition, we will require a normalization condition on the member functions akin to the cover condition of a crisp partition. This in turn will allow us to define the *norm* or size of a fuzzy cluster. For the sake of clarity, this condition will be introduced below.

B. Representations of a partition

1. *Matrix representation*

For any partition $P \in \Pi(S)$ we define the (*crisp*) *Adjacency Matrix* of P, $\mathbf{A}(P)$, as the $N \times N$ matrix given

by

$$\mathbf{A}(P)^{\alpha}{}_{\beta} \equiv \delta^{P(\alpha)}{}_{P(\beta)} \in \{0,1\} \tag{15}$$

where here δ is the Kronecker delta.

The adjacency matrix determines uniquely a partition as it is a representation of the corresponding equivalence relation. We will require that any general adjacency matrix of a partition P to be a real, positive semidefinite matrix satisfying the following general conditions

$$\mathbf{A}^{\alpha\alpha} \neq 0 \tag{16a}$$
$$\mathbf{A}^{\alpha\beta} = \mathbf{A}^{\beta\alpha} \tag{16b}$$
$$\mathbf{A}^{\alpha\beta} = \mathbf{A}^{\alpha\gamma}\mathbf{A}_{\gamma}{}^{\beta}, \tag{16c}$$

which correspond to the conditions of reflexivity, symmetry and transitivity, respectively, of the equivalence relation. An alternative formulation of the transitivity condition is in terms of max-min condition

$$\mathbf{A}^{\alpha\beta} = \sup_{\gamma}\{\min(\mathbf{A}^{\alpha\gamma}, \mathbf{A}^{\gamma\beta})\}$$

which we will not pursue in this work, however. For a continous (fuzzy) partition, we define the adjacency matrix as

$$\mathbf{A}(P)^{\alpha}{}_{\beta} \equiv \boldsymbol{\delta}^{\alpha\, i}\boldsymbol{\delta}^{\dagger}{}_{i}^{\beta} \tag{17}$$

which coincides with the crisp case (15)when

$$\boldsymbol{\delta}^{\alpha\, i} = \delta^{P(\alpha)\, i}. \tag{18}$$

For the particular representation of Eq.(15), the transitivity condition (16c) entails the following property of $\mathbf{A}(P)$

$$\left(\mathbf{A}^2\right)_{\alpha\,\beta} = (\mathbf{B}\,\mathbf{A})_{\alpha\beta}\,, \tag{19a}$$

where

$$\mathbf{B}_{\alpha\,\beta} \equiv n_{P(\alpha)}\,\delta_{\alpha\,\beta}, \tag{19b}$$

and $n_{P(\alpha)} = \|P_{P(\alpha)}\|$ is the number of elements in cluster $P(\alpha)$.

An alternative unique representation is given by the matrix

$$\mathbf{a}^{\alpha\ i} \equiv \delta^{P(\alpha)\ i} \tag{20a}$$
$$\mathbf{a}^{\dagger}{}_{i\ \alpha} \equiv \delta_{i\ P(\alpha)} \tag{20b}$$

where $\mathbf{a}^{\dagger}$ is the adjoint of $\mathbf{a}$, which in the present case coincides with its transpose. This is an $N\times\mathcal{K}$ matrix where each column vector specifies the elements belonging to that cluster. We will call this matrix the *pseudo-Ladder* matrix. By construction we have

$$\mathbf{A}(P) = \mathbf{a}\mathbf{a}^{\dagger} \tag{21}$$

We introduce the occupation number operator $\hat{\mathbf{N}}$ as

$$\hat{\mathbf{N}} \equiv \mathbf{a}^{\dagger}\mathbf{a}. \tag{22}$$

This is a diagonal matrix describing the number of elements in each cluster

$$\hat{\mathbf{N}}_{i}{}^{j} = n_i\,\delta_{i}{}^{j}. \tag{23}$$

Furthermore, $\mathbf{a}^{\dagger}$ corresponds to the matrix representation of the canonical mapping in a sense that will become clear later, once a graph is introduced. Thus, for any matrix $\boldsymbol{\mu}$ in the space of elements of S, its mapping in the space of clusters will be given by

$$\mathbf{M} = \mathbf{a}^{\dagger}\,\boldsymbol{\mu}\,\mathbf{a}. \tag{24}$$

A particular interesting case is when $\boldsymbol{\mu} = \mathbf{1}_{N\times N}$ is the identity. This would correspond to a similarity matrix describing only self-loops or *masses*. Its mapping into cluster space is $\hat{\mathbf{N}}$, which provides an alternative, graphical view of the number operator.

We can express the pseudo-ladder operator as a column matrix of N row vectors $\overrightarrow{a}^{\dagger}_{\alpha}$ specifying the cluster where element α belongs to, or alternatively, as a row matrix of $\mathcal{K}$ column vectors $\overrightarrow{a}_{i}$ specifying the elements belonging to cluster i, i.e.,

$$\mathbf{a} = \begin{bmatrix} \vdots \\ \overrightarrow{a}^{\dagger}_{\alpha} \\ \vdots \end{bmatrix} = \left[\cdots \overrightarrow{a}_{i} \cdots\right] \tag{25}$$

with $(\overrightarrow{a}_{\alpha})_i \equiv \delta_{P(\alpha)\,i}$ and $(\overrightarrow{a}_{i})^{\alpha} \equiv \delta_{i}^{P(\alpha)}$. From relations (21) and (22) we have by construction that these vectors are eigenvectors of the adjacency matrix and the number operator

$$\hat{\mathbf{N}}\,\overrightarrow{a}_{\alpha} = n_{P(\alpha)}\,\overrightarrow{a}_{\alpha} \tag{26a}$$
$$\mathbf{A}(P)\,\overrightarrow{a}_{i} = n_i\,\overrightarrow{a}_{i}. \tag{26b}$$

Thus, vectors $\overrightarrow{a}_{\alpha}$ and $\overrightarrow{a}_{i}$ are (up to a multiplication by a scalar) the right and left singular eigenvectors of $\mathbf{a}$, respectively, with singular value $\sqrt{n_i}$

$$\mathbf{a}\,\overrightarrow{a}_{\alpha} = \sqrt{n_{P(\alpha)}}\,\hat{a}_{P(\alpha)}$$
$$\mathbf{a}^{\dagger}\,\hat{a}_{P(\alpha)} = \sqrt{n_{P(\alpha)}}\,\overrightarrow{a}_{\alpha} \tag{27}$$

where $\hat{a}_i \equiv \frac{\overrightarrow{a}_i}{\sqrt{n_i}}$ is the normalized cluster vector, such that

$$g_{ij} \equiv \overrightarrow{a}_{i}\cdot\overrightarrow{a}_{j} = n_i\,\delta_{ij} = \hat{\mathbf{N}}_{ij} \tag{28a}$$
$$g_{\alpha\beta} \equiv \overrightarrow{a}_{\alpha}\cdot\overrightarrow{a}_{\beta} = \delta_{P(\alpha)P(\beta)} = \mathbf{A}_{\alpha\beta}. \tag{28b}$$

Relations (28) simply state that $\mathbf{A}$, and therefore $\hat{\mathbf{N}}$ too, is a Gram matrix, which follows directly from the requirement of being a positive semidefinite matrix. We will later see that the tensor products $\overrightarrow{a}_{\alpha}\otimes\overrightarrow{a}_{\beta}$ can indeed be interpreted as ladder operators of a particular Lie algebra. This justifies the name of pseudo-ladder operator for $\mathbf{a}$.

2. Geometric Algebra of partitions

Q 1. *A fundamental question is the representation of the intersection operation.*

Given two partitions $\mathbf{a}_P = \left[\cdots a_P^i \cdots\right]$ and $\mathbf{a}_Q = \left[\cdots a_Q^i \cdots\right]$, what is their relation to the intersection of P and Q, denoted here as $\mathbf{a}_{P\wedge Q} \equiv [\cdots \tilde{u}_i \cdots]$? We conjecture that $\mathbf{a}_{P\wedge Q}$ is basically but a orthogonalization of the combined set of cluster vectors of P and Q, followed by a rotation in a N-dimensional space. The latter is the underlying vectorial space spanned by the element unit vectors $\{e_\alpha\}$, with $e_\alpha \cdot e_\beta = \delta_{\alpha\beta}$. It is assumed that $\tilde{u}_i^\alpha \in \{0,1\}$ and $\tilde{u}_i \cdot \tilde{u}_j = \|\tilde{u}_i\|^2\,\delta_{ij}$, with only the first $\mathcal{K}_{\mathcal{P}\wedge\mathcal{Q}}$ vectors different from zero.

More specifically, this conjecture can be formulated as follows. Let's define the matrix $\Lambda \equiv \left[\cdots a_P^{i_P} \cdots a_Q^{i_Q} \cdots\right]$, which combines all columns from $\mathbf{a}_P$ and $\mathbf{a}_Q$, and for which it is $det(\Lambda) = 0$ and $rank(\Lambda) \equiv r$. Its QR decomposition, $\Lambda = \mathbf{QR}$, provides the orthonormalization of its column vectors as given by the orthogonal matrix

$$\mathbf{Q} \equiv \left[\cdots \frac{u_i}{\|u_i\|} \cdots\right], \tag{29}$$

where $i = 1, \cdots, r$. Furthermore, let's consider the r-dimensional oriented volume

$$\mathcal{V} \equiv u_1 \wedge \cdots \wedge u_r \tag{30}$$

given by the outer product of the orthogonal system $\{u_i\}$, as defined in the Geometric Algebra [55] GA_S over the vectorial space spanned by $\{e_\alpha\}$ (e_α denotes the singleton cluster containing only element α). Then, the set $\{e_\alpha \cdot \mathcal{V}_r\}$, where $e_\alpha \cdot \mathcal{V}_r \equiv \langle e_\alpha \mathcal{V}_r \rangle_{r-1}$ is the inner product of GA_S, defines a set of homogeneous multivectors of grade $r-1$ for which

$$rank\{e_\alpha \cdot \mathcal{V}_r\} = \mathcal{K}_{P\wedge Q} \tag{31a}$$

and

$$\sum_{i_1<\cdots<i_r} \lambda^{i_1\cdots i_r}\, \tilde{u}_{i_1} \wedge \cdots \wedge \tilde{u}_{i_r} = \mathcal{V}_r, \tag{31b}$$

where $\lambda^{i_1\cdots i_r} \in \mathcal{R}$ and $i_j = 1, \cdots, \mathcal{K}_{P\wedge Q}$ ranges over the non-zero cluster vectors of the intersection partition $P \wedge Q$. In summary, our conjecture claims that

Conjecture 1. *the set of clusters of $P \wedge Q$, $\{\tilde{u}_i\}$, is the set of $\mathcal{K}_{P\wedge Q}$ positive-defined vectors that satisfy (31b).*

As an example, let's consider the partitions of 5 elements $P = 23|145$ and $Q = 24|135$, with the intersection given by $P \wedge Q = 15|2|3|4$. The combined cluster matrix is

$$\Lambda = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \tag{32}$$

which has $r \equiv rank(\Lambda) = 3$. A Gram-Schmidt orthogonalization of its columns yields

$$[\cdots u_i \cdots] = \begin{bmatrix} 0 & 1 & -2 \\ 1 & 0 & 3 \\ 1 & 0 & -3 \\ 0 & 1 & 4 \\ 0 & 1 & -2 \end{bmatrix}, \tag{33}$$

which span the oriented 3-volume

$$\begin{aligned} \mathcal{V}_3 &\equiv u_1 \wedge u_2 \wedge u_3 = \\ &= 6\,(\, e_1e_2e_3 + e_2e_3e_5 - e_1e_2e_4 - e_2e_4e_5 \\ &+ e_2e_3e_4 - e_1e_3e_4 - e_3e_4e_5\,). \end{aligned} \tag{34}$$

The corresponding set of bivectors

$$\begin{aligned} e_1 \cdot \mathcal{V}_3 &= 6(e_2e_3 - e_2e_4 - e_3e_4) \\ e_2 \cdot \mathcal{V}_3 &= 6(-e_1e_3 + e_1e_4 + e_3e_4 - e_4e_5 + e_3e_5) \\ e_3 \cdot \mathcal{V}_3 &= 6(e_1e_2 - e_2e_4 - e_2e_5 + e_1e_4 - e_4e_5) \\ e_4 \cdot \mathcal{V}_3 &= 6(-e_1e_2 + e_2e_3 + e_2e_5 - e_1e_3 + e_3e_5) \\ e_5 \cdot \mathcal{V}_3 &= e_1 \cdot \mathcal{V}_3 \end{aligned}$$

has $rank\{e_\alpha \cdot \mathcal{V}_3\} = 4 = \mathcal{K}_{P\wedge Q}$ and it is straightforward to check that the set of orthogonal non-negative, cluster vectors corresponding to $\mathbf{a}_{P\wedge Q}$

$$\begin{aligned} \tilde{u}_1 &\equiv e_1 + e_5 \\ \tilde{u}_2 &\equiv e_2 \\ \tilde{u}_3 &\equiv e_3 \\ \tilde{u}_4 &\equiv e_4 \end{aligned}$$

satisfies Eq. (31b) with $\lambda_{123} = \lambda_{234} = 6$ and $\lambda_{234} = \lambda_{134} = -6$.

This conjecture unearths an underlying geometric structure of the algebra of partitions undocumented so far in the literature. A partition of N elements is associated to an oriented K-dimensional volume in the 1st quadrant, spanned by its K cluster vectors ($K \leq N$).

However, the previous discussion fails to provide a clear link between a partition and its supposedly associated volume element and can be considered as a suggestion that representing a cluster as a direct sum of its elements's vectors is not the most useful approach. Rather, describing clusters as multivectors (blades) provides a clear formulation of the intersection of partitions in terms of Geometric Algebra.

The intersection of two clusters can also be expressed geometrically using their corresponding oriented volumes. Consider the two clusters $\mathbf{a}_P = e_1 + e_2 + e_3$ and $\mathbf{b}_Q = e_1 + e_2 + e_4$. We may choose an arbitrary orientation for each cluster's volume. $\mathcal{V}_{a_P} = e_1e_2e_3$ and $\mathcal{V}_{b_Q} = e_1e_2e_4$. To the intersection of $\mathbf{a}_P$ and $\mathbf{b}_Q$ we can then associate the volume

$$\mathcal{V}_{\mathbf{a}_P\wedge\mathbf{b}_Q} = \mathcal{V}_a^\dagger\, \mathcal{V}_b\, I_{\mathbf{ab}}^\dagger, \tag{37}$$

modulo a sign, and where $I_{\mathbf{ab}} \equiv e_1e_2e_3e_4$ is the unit volume spanned by $\mathbf{a}$ and $\mathbf{b}$, and $\mathcal{V}^\dagger$ is the operation of

reversing the order of the geometric products in $\mathcal{V}$. That is, basically the volume associated with the intersection of two clusters is the dual of the product of their volumes. The reversing of $\mathcal{V}_{\mathbf{a}_P}$ is simply a matter of choice of sign. Explicitly we have

$$\mathcal{V}_{\mathbf{a}_P \wedge \mathbf{b}_Q} = (e_3 e_2 e_1)\,(e_1 e_2 e_4)\,e_4 e_3 e_2 e_1 = e_2 e_1. \quad (38)$$

If reversing the order of the clusters we obtain a volume element with opposite sign

$$\mathcal{V}_{\mathbf{a}_P \wedge \mathbf{a}_Q} = (e_4 e_2 e_1)\,(e_1 e_2 e_3)\,e_4 e_3 e_2 e_1 = -e_2 e_1. \quad (39)$$

According to our definition, however, both are equivalent geometric descriptions of the intersection of two clusters and we have that

$$c_{\mathbf{ab}} \equiv \mathbf{a}_P \wedge \mathbf{a}_Q = e_1 + e_2. \quad (40)$$

We can see now how to formulate the operation of intersection of partitions in the language of Geometric Algebra:

Lema III.1. **I**) *To any partition P with $\mathcal{K}_Q$ clusters we assign an oriented volume element given by*

$$\hat{\mathcal{V}}_P \equiv \bigoplus_{i=1,\ldots,\mathcal{K}_P} \mathcal{V}_{P_i}, \quad (41a)$$

where $\mathcal{V}_{P_i}$ is the oriented volume element associated to cluster P_i of partition P that is obtained from the geometric product of P_i's element vectors.

II) *Be partitions P and Q with $\mathcal{K}_Q$ and $\mathcal{K}_Q$ clusters, respectively. The oriented volume associated to the intersection $P \wedge Q$ is given by*

$$\hat{\mathcal{V}}_{P\wedge Q} = \hat{\mathcal{V}}_P \triangle \hat{\mathcal{V}}_Q \equiv \bigoplus_{\substack{i=1,\ldots,\mathcal{K}_P \\ j=1,\ldots,\mathcal{K}_Q}} \mathcal{V}^\dagger_{P_i} \mathcal{V}_{Q_j} I^\dagger_{P_i Q_j} \quad (41b)$$

The *triangle product* of oriented volume elements defined in Eqs.(41) is our, so far simplest, Geometric Algebra equivalent of the operation of intersection of two partitions. To this point (Fri., June 7th 2013) we don't know if the triangle product can be expressed using only standard operations of Geometric Algebra on the volumes $\mathcal{V}_P$ and $\mathcal{V}_Q$.

(Update, Wed. Nov. 19th 2014) It seems to me we are dealing with a pair of dual characterizations of a cluster. On one hand, when we define the n-volume $\mathcal{V}_P$, we treat the clusters as a direct sum of element vectors, $a_{P_i} \equiv \oplus_j e_j$. On the other hand, in (41) a cluster is defined as a blade of the corresponding geometric algebra, i.e., $a_{P_I} \equiv \wedge_j e_j$.

It is also easy to check that our conjecture of the association of an oriented volume to any partition is at least consistent, namely, it is satisfied when we calculate the intersection of a partition P with itself. Furthermore, in such a case the only characteristic n-volume that appears is precisely the one spanned by its cluster *vectors*. In other words, this suggests that we can indeed assign

$$\mathcal{V}_P \equiv \wedge_i a_{P_i}. \quad (42)$$

It is still not clear, though, what the role of $\mathcal{V}_r$ in (31b) is.

Nevertheless, the idea of a dual description might be a useful lead in this matter.

Conjecture 2. *For any partition P with cluster vectors a_{P_i}, $i \in [1, \mathcal{K}]$, its associated oriented volume $\mathcal{V}_P$ and the clusters' blades $\mathcal{V}_{a_{P_i}}$ are related by*

$$\mathcal{V}_P = \mathcal{V}_{a_{P_1}}\, a_{P_1} \cdots \mathcal{V}_{a_{P_\mathcal{K}}}\, a_{P_\mathcal{K}}\, I, \quad (43)$$

where I is the unit volume spanned by all element vectors e_i.

This relation, if true, would justify the "nicer" geometric description of a partition as a multivector given by the direct sum of its clusters blades, i.e.,

$$P := \bigoplus_{i=1,\cdots,\mathcal{K}_P} \mathcal{V}_{a_{P_i}}. \quad (41a)$$

Nicer here is meant in the sense that (41a) has a more direct geometric interpretation as a direct sum of elementary k-parallelotopes (each a unit pseudoscalar) of varying dimension ($k = 1, \cdots, n_\mathcal{K}$) of unit (hyper-)volume and all sharing the origin as the only single point in contact between them. Thus all singletons are unit, oriented segments perpendicular to each other and the rest of clusters; all 2-elements clusters are oriented squares of unit area perpendicular to every singleton and the rest of clusters; all 3-elements clusters are oriented cubes of unit volume with faces that are perpendicular to all 2-elements clusters and all singletons, as well as to those of other 3-elements clusters; etc. In all cases, the hyper-volume is defined as $|\mathcal{V}_{a_{P_i}}| \equiv \mathcal{V}_{a_{P_i}} \mathcal{V}^\dagger_{a_{P_i}}$.

As an example, let's consider the partition $P = 123|45|6$. It is

$$\begin{aligned}
\mathcal{V}_{a_{P_1}} &\equiv e_1\, e_2\, e_3 \\
\mathcal{V}_{a_{P_2}} &\equiv e_4\, e_5 \\
\mathcal{V}_{a_{P_3}} &\equiv e_6 \\
\mathcal{V}_P &\equiv a_{P_1} \wedge a_{P_2} \wedge a_{P_3} = \\
&= e_1\, e_4\, e_6 + e_1\, e_5\, e_6 + e_2\, e_4\, e_6 + e_2\, e_5\, e_6 + \\
&+ e_3\, e_4\, e_6 + e_3\, e_5\, e_6
\end{aligned}$$

from which we obtain

$$\begin{aligned}
\mathcal{V}_P\, I^\dagger &= \mathcal{V}_P\, e_6\, e_5\, e_4\, e_3\, e_2\, e_1 = \\
&= e_5\, e_3\, e_2 - e_4\, e_3\, e_2 + e_5\, e_2\, e_1 - e_4\, e_2\, e_1 - \\
&- e_5\, e_3\, e_1 + e_4\, e_3\, e_1 \\
\mathcal{V}_{a_{P_1}}\, a_{P_1} &= e_2\, e_3 - e_1\, e_3 + e_1\, e_2 = \mathcal{V}_{a_{P_1}} \cdot a_{P_1} \\
\mathcal{V}_{a_{P_2}}\, a_{P_2} &= -e_5 + e_4 \\
\mathcal{V}_{a_{P_3}}\, a_{P_3} &= 1
\end{aligned}$$

yielding

$$\begin{aligned}
\mathcal{V}_P\, I^\dagger &= -e_5\, \mathcal{V}_{a_{P_1}}\, a_{P_1} + e_4\, \mathcal{V}_{a_{P_1}}\, a_{P_1} = \\
&= \mathcal{V}_{a_{P_2}}\, a_{P_2}\, \mathcal{V}_{a_{P_1}}\, a_{P_1} = \\
&= \mathcal{V}_{a_{P_1}}\, a_{P_1}\, \mathcal{V}_{a_{P_2}}\, a_{P_2}\, \mathcal{V}_{a_{P_3}}\, a_{P_3}
\end{aligned}$$

which alternatively can be expressed as

$$\mathcal{V}_P\, I^\dagger \;=\; \left(\mathcal{V}_{a_{P_1}}\cdot a_{P_1}\right)\,\left(\mathcal{V}_{a_{P_2}}\cdot a_{P_2}\right)\,\left(\mathcal{V}_{a_{P_3}}\cdot a_{P_3}\right).$$

Q 2 (Feb. 4th 2013)**.** *What is the relation to a Plücker embedding?*

I just came across the concept of random partitions and the Plücker embedding of a Grassmanian associated with a Ferres diagram λ (drawn *in the Russian way* where the vertex of the diagram lies at the origin and the largest cluster lies along the main diagonal of the first quadrant, and the rest are piled atop of it; the partition grows this way along the second diagonal). This seems to be somewhat of a recurrent concept relating to ramdom matrices, quatum crystals and spin chains among other topics [56, 57]. It's not clear to me, however, to what extent the oriented volume we introduced here is analogous to that given by the Plücker embedding. The latter assigns an orthonormal vector to each cluster in an N-dimensional vectorial space along a direction that depends on the size of the cluster. Furthermore, there is the *infinite* contribution of the remaining non-occupied clusters beyond the last smallest actual cluster of a given partition. These represent the so-called *fermi sea*, and each actual cluster represents either an *excited* (above the fermi sea) particle, equivalently, a hole in that sea, or a particle within it. Furthermore, the projection of each cluster height onto dots on the horizontal axis provides a reprentation of these particles as the eigenvalues of random matrices. The study of correlations and probability distributions of these dots links to random matrices and to the problem of determining probability distributions on the space of all (integer) partitions of N. The (infinite) exterior product of the vectors thus defined determines an oriented volume or multivector $\mathcal{V}_\lambda$ associated to any given Ferres diagram. In our case, the oriented volume $\mathcal{V}_P$ is also obtained as an exterior product of orthonormal cluster vectors in an N-dimensional space, but it is a *finite*, as opposed to infinite product and each vector *depends on the particular content of the corresponding cluster*. Hence, the mapping $\lambda(P)\mapsto\mathcal{V}_{\lambda(P)}$ is not well defined as it assigns more than one oriented volume to the same Ferres diagram λ.

I suspect the difference in having or not an infite number of vectors represents just that we work in a *micro-canonical* ensemble description while these other representations work in a *grand-canonical* ensemble. More important is thus the difference in the way of labeling and assigning vectors to each cluster. However, maybe we can learn and adapt something from those cases, like trying to come with a Fock space and its ground state in a similar way. The Plancherel measure is giving by $\mathfrak{M}_{Planch}(\lambda)=\frac{(\dim\lambda)^2}{|\lambda!}$.

It bothers me, though, that they define the energy operator L_0, and thus, the energy, as $L_0\cdot\mathcal{V}_\lambda=|\lambda|\mathcal{V}_\lambda$, where $|\lambda|=N$.

NOTE: (Fri. Aug. 5th 2016)

It's easier to understand the Plancherel measure as defined in wikipedia: *Let G be a finite group and denote by G^* the set of its irrep. The corresponding Plancherel measure over the set G^* is defined as*

$$\mu(\pi)\;=\;\frac{(dim\,\pi)^2}{|G|}\tag{44}$$

where $\pi\in G^*$ and $dim\,\pi$ is the dimension (or order) of the irrep π. For instance, as S_3 has 3 conjugacy classes, for its regular representation, besides the totally symmetric and totally antisymetric invariant submodules of order 1, there is another one of order $d=2$ ($3!=1^2+1^2+d^2$). Thus it decomposes into 3 irreps with characters $\vec{\chi}^{(1)}=(1,1,1,1,1,1)$, $\vec{\chi}^{(2)}=(1,-1,-1,-1,1,1)$, $\vec{\chi}^{(3)}=(2,0,0,0,-1,-1)$.

Furthermore, under Sage, the function `plancherel_measure()` returns the probability under the Plancherel probability measure on partitions of the same size. Thus

```
Partitions(2).list()
   [[2],[1,1]]
Partitions([2]).plancherel_measure()
   1/2
Partitions(3).list()
   [[3], [2, 1], [1, 1, 1]]
[mu.plancherel_measure() for mu in Partitions(3)]
   [1/6, 2/3, 1/6]
Partitions(4).list()
   [[4], [3, 1], [2, 2], [2, 1, 1], [1, 1, 1, 1]]
[mu.plancherel_measure() for mu in Partitions(4)]
   [1/24, 3/8, 1/6, 3/8, 1/24]
```

Note that there are five partitions of the set $S=1,2,3$ namely

```
[1,2,3]
[[1,2],[3]] *
[[1],[2],[3]]
```

(*there are 3 partitions with this block-size distribution, namely (2,1)). Therefore, the Plancherel measure is not a measure in the Hasse lattice of partitions of a set.

On the contrary, and what makes it more interesting, is that the Plancherel measure exemplifies the link between the irreps of S_n and the partitions of n. For instance, for S_3 and $\pi^{(2)}\equiv\vec{\chi}^{(2)}$, whe have $|S_3|=6$ and $dim\,\pi=2$, therefore $\mu(\pi^{(2)})=2^2/6=2/3$. Furthermore, $\pi^{(2)}$ is associated with the Young diagram corresponding to the integer partition with block sizes $(2,1)$.

Remark: The number $p(n)$ of partitions of n is given by the generating function

$$Z_p\;\equiv\;\sum_{n=0}^{\infty}p(n)\,q^n\;=\;\prod_{j=1}^{\infty}\frac{1}{1-q^j},\;\;|q|<1.\tag{45}$$

3. *Vector representation*

As (15) and (20) are matrix representations of a partition P, one is tempted to define as well a corresponding

vectorial representation of it. In the following, we will use Dirac's bra-ket notation for describing vectors and their transpose. Let $\mid k\rangle$ represent the k-th cluster and lets assume $\{\mid k\rangle\}$ is an orthonormal system of vectors, i.e.

$$\langle k' \mid k\rangle \;=\; \delta_{kk'}. \tag{46}$$

Analogously, let $\mid \alpha\rangle$ represent a node (*state*) of the set of elements (*system*) we want to describe. Mimicking the usual structure of linear algebras, to any particular realization of the partition P we can ascribe a ket as

$$\mid P\rangle \;=\; \sum_{\alpha i} \mathbf{a}^{\alpha}{}_{i} \mid \alpha\rangle \mid i\rangle \tag{47}$$

Up to this point, (47) is just an ad-hoc representation equivalent to (20). In order to have a complete sense we would have to prove that the set of all vectors $\mid P\rangle$ really span a vectorial space, i.e., we would need to define what is meant by the *sum* of two partitions $\mid P\rangle + \mid Q\rangle$, as well as the *multiplication* by a scalar $\lambda \mid P\rangle$. While, mathematically, both operations are well defined their interpretation in the present context in not clear. We will postpone this discussion for later however. The aim, here, is rather to explore the possibility of an alternative description of a partition on a graph in terms of the usual formalism of Statistical Mechanics.

An operator (matrix) $\boldsymbol{\omega}$ acting only on the nodes means

$$\boldsymbol{\omega}\left(\mid \alpha\rangle \mid i\rangle\right) \;=\; \left(\boldsymbol{\omega} \mid \alpha\rangle\right) \mid i\rangle, \tag{48}$$

and it's expression in the combined space (tensorial product) $\mid \alpha\rangle \mid i\rangle$ is given by

$$\boldsymbol{\omega} \;=\; \sum_{\alpha\beta ii'} \boldsymbol{\omega}_{\alpha\beta} \mid \alpha\rangle \mid i\rangle\langle i' \mid \langle\beta \mid \tag{49}$$

For any given set S and a partition of it P, we will consider that the $\mid \alpha\rangle$'s in node space, as well as the $\mid k\rangle$'s in cluster space, each set constitute a complete orthonormal set of vectors in their respective spaces. Thus, the matrix

$$\mathcal{I} \;\equiv\; \sum_{\alpha} \mid \alpha\rangle\langle\alpha \mid \tag{50}$$

satisfies

$$\mathcal{I} \mid \beta\rangle \;=\; \mid \beta\rangle \tag{51}$$

$$\langle\alpha \mid \mathcal{I} \mid \beta\rangle \;=\; \langle\alpha \mid \beta\rangle \;=\; \delta_{\alpha\beta}. \tag{52}$$

With the above definitions the norm of $\mid P\rangle$ is

$$\begin{aligned}\langle P \mid P\rangle \;&=\; \sum_{\alpha\beta ij} \langle\alpha \mid \langle i \mid \mathbf{a}_{i}{}^{\alpha}\mathbf{a}^{\beta}{}_{j} \mid \beta\rangle \mid j\rangle \\ &=\; \sum_{\alpha i} \mathbf{a}_{i}{}^{\alpha}\mathbf{a}^{\alpha}{}_{i} \\ &=\; Tr\left(\mathbf{A}\right) \;=\; Tr\left(\hat{\mathbf{N}}\right) \;=\; N.\end{aligned} \tag{53}$$

It will be convenient to have $\langle P \mid P\rangle = 1$ for which we will normalize the $\mathbf{a}$ by $\sqrt{N}$, i.e., we will define the *state* vector representing a partition P as

$$\mid P\rangle \;=\; \sum_{\alpha i} \tilde{\mathbf{a}}^{\alpha}{}_{i} \mid \alpha\rangle \mid i\rangle, \tag{54}$$

where

$$\tilde{\mathbf{a}}^{\alpha}{}_{i} \;\equiv\; \frac{1}{\sqrt{N}}\mathbf{a} \;=\; \frac{1}{\sqrt{N}}\delta^{P(\alpha)}_{\;\;i}. \tag{55}$$

C. Distribution (density) matrix approach

Following [36], "a pure ensemble by definition is a collection of physical systems such that every member is characterized by the same ket $\mid i\rangle := \overrightarrow{a}_{i}$. In contrast, in a mixed ensemble, a fraction of the members with relative population f_1 is characterized by $\mid 1\rangle$, some other fraction with relative population f_2, by $\mid 2\rangle$, and so on. Roughly speaking, a mixed ensemble can be viewed as a mixture of pure ensembles, just as the name suggests. The fractional populations are constrained to satisfy the normalization condition $\sum_i f_i = 1$".

The density operator is then given by

$$\rho \;\equiv\; \sum_{i} f_i \mid i\rangle\langle i \mid := \boldsymbol{a}\boldsymbol{f}\boldsymbol{a}^{\dagger}, \tag{56}$$

where the diagonal matrix $\boldsymbol{f} = diag\,(f_1, f_2, \dots)$ describes the distribution of relative (cluster) populations.

IV. SYMMETRIES OF PARTITIONS

A. The fundamental relation of partitions

The singular value decomposition of $\mathbf{a}$ is given by

$$\mathbf{a} \;=\; \mathbf{U}\hat{\mathbf{a}}\mathbf{V}^{\dagger}, \tag{57}$$

where $\hat{\mathbf{a}}$ is a $N\times\mathcal{K}$ diagonal matrix, $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices of dimensions $N\times N$ and $\mathcal{K}\times\mathcal{K}$, respectively, and are given by

$$\mathbf{V} \;\equiv\; \left[\cdots \overrightarrow{a}_{\alpha} \cdots\right] \;=\; \mathbf{1}_{\mathcal{K}} \tag{58}$$

$$\mathbf{U} \;\equiv\; \left[\hat{a}_1 \cdots \hat{a}_{\mathcal{K}}\, \hat{b}_1 \cdots \hat{b}_{N-\mathcal{K}}\right], \tag{59}$$

where $\mathbf{V}$ is build by choosing one single archetypal element of each cluster, and where the set of orthonormal vectors $\{\hat{b}_i \equiv \hat{a}_{\mathcal{K}+i}\}_{i=1,\cdots,N-\mathcal{K}}$ spans the null-space of $\mathbf{a}^{\dagger}$. For boolean partitions, $\mathbf{V}$ is simply the $\mathcal{K}\times\mathcal{K}$ identity matrix $\mathbf{1}_{\mathcal{K}}$. In terms of $\hat{\mathbf{a}}$, the adjacency matrix of P is $\mathbf{A}(P) = \mathbf{U}\hat{\mathbf{a}}\hat{\mathbf{a}}^{\dagger}\mathbf{U}^{\dagger}$, being thus $\mathbf{U}$ an isometry that diagonalizes $\mathbf{A}(P)$. Furthermore, and with the appropriate

reordering of rows and columns, it is

$$\hat{\mathbf{a}} = \begin{bmatrix} \sqrt{\hat{\mathbf{N}}} \\ \mathbf{0} \end{bmatrix} \tag{60a}$$

$$\hat{\mathbf{a}}^\dagger \hat{\mathbf{a}} = \hat{\mathbf{N}} \tag{60b}$$

$$\hat{\mathbf{a}} \hat{\mathbf{a}}^\dagger = \begin{bmatrix} \hat{\mathbf{N}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{60c}$$

The previous definition of $\mathbf{a}$ as a $N \times \mathcal{K}$ matrix seems a natural one and has been used in the literature before [54], but limiting the number of columns to $\mathcal{K}$ complicates unnecessarily the algebra and expressions involved. Thus, we will heretofore redefine $\mathbf{a}$, and therefore also $\hat{\mathbf{a}}$, as an $N \times N$ matrix with the first $\mathcal{K}$ columns corresponding to non-vanishing cluster vectors $\overrightarrow{a}_i$, and the last $N - \mathcal{K}$ columns containing only zeros. This means $\hat{\mathbf{N}}$ is also an $N \times N$ matrix, with only the first $\mathcal{K}$ diagonal elements different from zero. With this redefinition, relations (60) become

$$\hat{\mathbf{a}} = \sqrt{\hat{\mathbf{N}}} \tag{61a}$$

$$\hat{\mathbf{N}} = \mathbf{a}^\dagger \mathbf{a} = \hat{\mathbf{a}}^\dagger \hat{\mathbf{a}} = \hat{\mathbf{a}} \hat{\mathbf{a}}^\dagger = \hat{\mathbf{a}}^2 \tag{61b}$$

$$\hat{\mathbf{N}} = \begin{bmatrix} n_1 & & & & \\ & \ddots & & & \\ & & n_{\mathcal{K}} & & \\ & & & \mathbf{0} & \\ & & & & \ddots \end{bmatrix}. \tag{61c}$$

This allows us to write the fundamental relation of the matrix representation of a partition, namely

$$\mathbf{A}(P) = \mathbf{U} \hat{\mathbf{N}} \mathbf{U}^\dagger. \tag{62}$$

Thus, $\mathbf{U}$, which diagonalizes $\mathbf{A}(P)$, provides a mapping between the cluster-based representation and the element-based one. Furthermore, Eq.(62) conveys the fundamental idea of a partition as the rotation of a given distribution of cluster sizes $\{n_i\}$, with $N = \sum_i n_i$, such that $\mathbf{A}$ is a *proper* adjacency matrix, i.e., satisfies conditions (16). The fundamental relation (62) is a direct consequence of $\mathbf{A}$ being a real symmetric (16b) matrix. As such, it can always be diagonalized with the similarity transformation being an orthogonal matrix. From the above derivation, however, we also see that $\mathbf{U}$ has determinant ± 1. Hereafter, we will consider only those rotations with determinant $+1$, i.e., $\mathbf{U} \in SO(N)$.

For the general case of continuous partitions, all matrices are $N \times N$ dimensional and it is $\mathbf{V} \neq \mathbf{1}_N$. The fundamental relation becomes then

$$\mathbf{A} = \mathbf{R} \mathbf{N} \mathbf{R}^\dagger \tag{63a}$$

$$\mathbf{N} = \mathbf{V} \hat{\mathbf{N}} \mathbf{V}, \tag{63b}$$

where

$$\hat{\mathbf{N}} \equiv \hat{\mathbf{a}}^\dagger \hat{\mathbf{a}} \tag{63c}$$

$$\mathbf{R} = \mathbf{U} \mathbf{V}^\dagger, \tag{63d}$$

and $\hat{\mathbf{a}}$ is the diagonal matrix of singular values of $\mathbf{a}$, as in the case of boolean partitions. Thus the fundamental relation still holds for continuous or fuzzy partitions. The entries of $\mathbf{A}$ and $\mathbf{N}$ are no longer integers, however, but we will still impose the normalization condition of $Tr(\mathbf{A}) = Tr(\mathbf{N}) = N$.

B. Symmetries of partitions

Q 3. *An interesting question would be to develop in detail what conditions (16) entail for any partition rotation* $\mathbf{U}$.

For the particular case of a crisp partition, the transitivity condition entails

$$\hat{\mathbf{N}}^2 = \mathbf{U}^\dagger \mathbf{B} \mathbf{U} \hat{\mathbf{N}}. \tag{64}$$

However, we do not know whether this is valid in the general case of a fuzzy partition.

An additional property of $\mathbf{U}$ comes from its definition Eq.(59). We can write $\mathbf{U} = \hat{\mathfrak{a}} + \hat{\mathfrak{b}}$, where

$$\hat{\mathfrak{a}} = [\hat{a}_1, \ldots, \hat{a}_{\mathcal{K}}, 0 \ \ldots, 0] \tag{65a}$$

$$\hat{\mathfrak{b}} = [0 \ , \ldots, 0 \ , \hat{b}_1, \ldots, \hat{b}_{N-\mathcal{K}}] \tag{65b}$$

are two $N \times N$ matrices with their columns satisfying

$$\hat{a}_i \cdot \overrightarrow{1} = \sqrt{n_i} \tag{66a}$$

$$\hat{a}_i^\alpha \geq 0$$

$$\hat{b}_i \cdot \overrightarrow{1} = 0 \tag{66b}$$

$$\mathfrak{a}^\dagger \mathfrak{b} = \mathfrak{b}^\dagger \mathfrak{a} = 0. \tag{66c}$$

Matrix $\hat{\mathfrak{a}}$ differs from $\mathbf{a}$, as the former is obtained from the latter by normalizing all non-zero column vectors, and is also different from matrix $\hat{\mathbf{a}}$, which is a diagonal matrix containing the singular values of $\mathbf{a}$. In terms of matrices (65), the fundamental relation reads as

$$\mathbf{A} = \hat{\mathfrak{a}} \hat{\mathbf{N}} \hat{\mathfrak{a}}^\dagger, \tag{67}$$

where the orthogonality property (66c) has been used.

Relations (66b) and (66c) suggest that we can interpret $\mathfrak{b}$ as the matrix of *holes* $\hat{b}_j$. Consider a *bottom-up* clustering process which starts from partition $\bar{0} = 1|2|3|\cdots|\mathcal{K}$ ($\mathcal{K} = N$) and from there proceeds to find the optimal partition. At its first step, when merging any two singleton clusters, say $e^1 + e^2$ leading to partition $P = 12|3|4|\cdots$, we reduce $\mathcal{K}$ by 1. We can see this, however, as element 2 merging into the cluster of element 1, and leaving a *hole* where cluster e^2 was, and we could write $P = 12| \cdot |3|4|\cdots$. For instance, we can think of partition $P = 123|45$ as $P = 123| \cdot | \cdot |45| \cdot | = 123|45| \cdot | \cdot | \cdot |$. The orthonormal basis $\{\hat{b}_j\}$ represents those *hole* clusters, and relation Eq.(66b) the fact that they contain no element.

The composition with $\mathbf{U}$ of any rotation $\mathbf{R}$ that leaves invariant $\mathfrak{a}$ yields a new rotation $\mathbf{U}'$ that corresponds

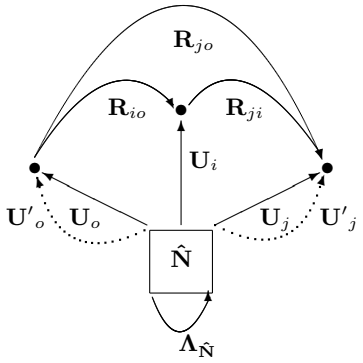


FIG. 3: Symmetries and transformations of partitions. To any given cluster-size distribution $\hat{\mathbf{N}}$ there corresponds several equivalence relations in S (filled circles; labels $\mathbf{A}_i$ not shown) which can be obtained from their corresponding partition rotation $\mathbf{U}_i$ as $\mathbf{A}_i = \mathbf{U}_i\hat{\mathbf{N}}\mathbf{U}_i^\dagger$. By choosing a reference partition $\mathbf{U}_o$ all other partitions can be obtained by a corresponding relative rotation $\mathbf{R}_{io}$ ($\mathbf{U}_i = \mathbf{R}_{io}\mathbf{U}_o$). Equivalent rotations $\mathbf{U'}_o$ describe the same partition and give rise to symmetries of their equivalence relation $\mathbf{A}_o$ (e.g., $\mathbf{U'}_o\mathbf{U}_o^\dagger$, described by the closed loop, continuous-dotted lines starting at $\mathbf{A}_o$ and moving in a clockwise orientation), as well as of $\hat{\mathbf{N}}$. The set of all rotations $\mathbf{\Lambda}_{\hat{\mathbf{N}}}$ that commute with $\hat{\mathbf{N}}$ characterizes all possible partitions associated with that given cluster-size distribution: For any $\mathbf{\Lambda}_{\hat{\mathbf{N}}}$, the rotation $\mathbf{U}_i\mathbf{\Lambda}_{\hat{\mathbf{N}}}^\dagger$ is equivalent to $\mathbf{U}_i$. Examples of such a rotation are $\mathbf{\Lambda}_{\hat{\mathbf{N}}} = \mathbf{U'}_o^\dagger\mathbf{U}_o$ and $\mathbf{\Lambda}_{\hat{\mathbf{N}}} = \mathbf{U}_j^\dagger\mathbf{R}_{ji}\mathbf{U}_i$. With the composition of rotations, the relative rotations $\mathbf{R}_{ij}$ form a Lie group $\boldsymbol{R}_{\hat{\mathbf{N}}} \subset SO(N)$.

to the same partition, i.e., yields the same equivalence relation $\mathbf{A}$. For such a rotation $\mathbf{R}$ it is

$$\begin{aligned}\mathbf{U}' &= \mathbf{R}\,\mathbf{U} = \mathbf{R}\,\mathfrak{a} + \mathbf{R}\,\mathfrak{b} \\ &= \mathfrak{a} + \mathbf{R}\,\mathfrak{b} \equiv \hat{\mathfrak{a}}' + \hat{\mathfrak{b}}' \qquad (68)\\ \mathbf{A}' &= \mathbf{U}'\,\hat{\mathbf{N}}\,\mathbf{U}'^\dagger \\ &= \hat{\mathfrak{a}}'\,\hat{\mathbf{N}}\,\hat{\mathfrak{a}}'^\dagger \\ &= \mathbf{R}\,\mathfrak{a}\,\mathbf{N}\,\mathfrak{a}^\dagger\,\mathbf{R}^\dagger \\ &= \mathbf{R}\,\mathbf{A}\,\mathbf{R}^\dagger = \mathbf{A}. \qquad (69)\end{aligned}$$

Therefore, the set of rotations $\boldsymbol{R}_{\mathbf{A}} = \{\mathbf{R}_{\mathbf{A}}(\theta)\}$, with θ a finite set of continuous parameters, that leave invariant a given relation $\mathbf{A}$ forms a Lie group $\boldsymbol{R}_{\mathbf{A}} \subset SO(N)$ describing the degeneracy in the description of the partition $\mathbf{U}$.

We can summarize the above discussion by defining the following equivalence relation among rotational representations of partitions. We will say that two partition rotations $\mathbf{U}$ and $\mathbf{U}'$ are equivalent iif

$$[\mathbf{A}\,,\,\mathbf{R}] = 0 \quad , \quad \mathbf{R} \equiv \mathbf{U}'\,\mathbf{U}^\dagger, \qquad (70a)$$

with $[,]$ the commutator, or, equivalently, iif

$$[\hat{\mathbf{N}}\,,\,\mathbf{Q}] = 0 \quad , \quad \mathbf{Q} \equiv \mathbf{U}'^\dagger\,\mathbf{U}, \qquad (70b)$$

and both are related by an additional pre- or post-rotation as $\mathbf{U}' = \mathbf{U}\mathbf{Q}^\dagger = \mathbf{R}\mathbf{U}$.

Furthermore, any two matrices $\mathbf{A}_1$ $\mathbf{A}_2$ corresponding to the same cluster-size distribution $\hat{\mathbf{N}}$ are related through a rotation

$$\mathbf{A}_2 = \mathbf{R}_{21}\,\mathbf{A}_1 \quad , \quad \mathbf{R}_{21} = \mathbf{U}_2\,\mathbf{U}_1^\dagger. \qquad (71)$$

Therefore, we can describe all partitions corresponding to a given cluster-size distribution $\hat{\mathbf{N}}$ by choosing one one $\mathbf{U}_o$ as a reference and listing the relative rotations $\mathbf{R}_{io}$ giving rise to the rest (see Fig.3). For instance, all partitions of $N = 4$ elements into two clusters can be divided into two equivalence classes, one with 3 partitions and the other with 4, with respective cluster-size distributions $\hat{\mathbf{N}}_{2a} = diag(2,2,0,0)$ and $\hat{\mathbf{N}}_{2b} = diag(3,1,0,0)$. Within each class, consider two partition rotations $\mathbf{U}_i$ and $\mathbf{U}_j$, the relative rotation $\mathbf{R}_{ji} = \mathbf{U}_j\mathbf{U}_i^\dagger$ transforms the first into the second, i.e., $\mathbf{U}_j = \mathbf{R}_{ji}\mathbf{U}_i$. Therefore, choosing one as a reference, e.g., $P_o = 12|34 \leftrightarrow \mathbf{U}_o$, we can generate all the rest of partitions corresponding to $\hat{\mathbf{N}}_{2a}$ through relative rotations.

With the composition of rotations, the relative rotations $\mathbf{R}_{ij}$ form a Lie group $\boldsymbol{R}_{\hat{\mathbf{N}}} \subset SO(N)$. An as of yet open problem is to characterize this subgroup of relative rotations. The constraints come from the condition that $\mathbf{R}_{io}\mathbf{U}_o$ must yield a valid equivalence relation.

An alternative point of view, is to consider all rotations $\mathbf{\Lambda}_{\hat{\mathbf{N}}}$ that commute with $\hat{\mathbf{N}}$. For any such rotation and any partition rotation $\mathbf{U}_i$ compatible with the given cluster-size distribution, the rotation $\mathbf{U}_i\mathbf{\Lambda}_{\hat{\mathbf{N}}}^\dagger$ is equivalent to $\mathbf{U}_i$. Examples of such a rotation are $\mathbf{\Lambda}_{\hat{\mathbf{N}}} = \mathbf{U'}_o^\dagger\mathbf{U}_o$ and $\mathbf{\Lambda}_{\hat{\mathbf{N}}} = \mathbf{U}_j^\dagger\mathbf{R}_{ji}\mathbf{U}_i$ (see Fig.3).

Q 4. *An open question is then how to generate all partition rotations given the set of symmetry rotations* $\{\mathbf{\Lambda}_{\hat{\mathbf{N}}}\}$.

Within a given level $\mathcal{K}$ of the Hasse diagram we assume there exist a transformation $\mathbf{T}$ such that for two cluster-size distributions $\hat{\mathbf{N}}$ and $\tilde{\mathbf{N}}$ (e.g. $\mathcal{K} = 2$ and $\hat{\mathbf{N}}_{2a}$ and $\hat{\mathbf{N}}_{2b}$), we have

$$\tilde{\mathbf{N}} = \mathbf{T}\,\hat{\mathbf{N}} \quad , \quad Tr(\tilde{\mathbf{N}}) = Tr(\hat{\mathbf{N}}). \qquad (72)$$

Let us consider the case of a crisp partition and a symmetry $\tilde{\mathbf{\Lambda}}$ of $\tilde{\mathbf{N}}$. For this case, $\mathbf{T}$ must be a diagonal matrix. Then the transformation $\mathbf{S} \equiv \mathbf{T}^{-1}\tilde{\mathbf{\Lambda}}\mathbf{T}$ is as well a symmetry of $\tilde{\mathbf{N}}$.

Q 5. *Can we define similar transformations between two cluster-size distributions of different levels* $\mathcal{K}$ *and* $\mathcal{K}'$*?*

Lema IV.1. *(Tue. Jun. 7 2016)For any two cluster-size distribution* $\tilde{\mathbf{N}}$ *and* $\tilde{\mathbf{N}}'$ *there is a similarity transformation* $\mathbf{S}$ *between them, i.e.,* $\tilde{\mathbf{N}}' = \mathbf{S}\tilde{\mathbf{N}}\mathbf{S}^{-1}$.

Be $\mathbf{G}$ any stochastic matrix of dimension N, i.e., $\sum_i G_{ij} = 1$. Then for any partition $|n\rangle = (n_1, n_2, \cdots, n_N)$ of N, such that $N = n_1 + \cdots n_N$, where not all n_i are zero, $|n'\rangle = \mathbf{G}|n\rangle$ is as well a partition of N. This induces a similarity transformation Γ of $\tilde{\mathbf{N}}$, for one cluster-size distribution to another arbitrary one. We have already seen that there are rotations $\mathbf{R}$ that can bring us from one partition to any other with the same cluster-size distribution $\tilde{\mathbf{N}}$. Thus, $\mathbf{S} \equiv \mathbf{R}\boldsymbol{\Gamma}$ is such a similarity transformation. □

Conjecture 3. *G is given by $G_{|n'\rangle|n\rangle} = \gamma_{|n\rangle}|n'\rangle\langle n|$, where $\gamma_{|n\rangle} = |\langle n|n\rangle|^{-1}$. Moreover, for any given N, the collection of its partitions $|n\rangle$ forms a groupoid with the G's as ismorphisms.*

Wrong! These do give rise to such a groupoid but, in general, are not invertible nor satisfy that their columns add up to 1. Forcing the latter yields G for which the components are not necessarily positive, i.e., are not stochastic matrices. Furthermore, the groupoid contains morphisms that may not be given by the tensor product of vectors $|n\rangle$, e.g, for $N = 3, |2\rangle \equiv |2\,1\,0\rangle, |1\rangle \equiv |1\,,1\,1\rangle$, we have the alternative morphism between $|1\rangle$ and $|2\rangle$ given by $G_{12} = 4^{-1}\begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 4 \end{pmatrix}$ and $G_{21} = 4^{-1}\begin{pmatrix} 8 & 0 & 0 \\ -4 & 8 & 0 \\ 0 & -4 & 4 \end{pmatrix} = (G_{12})^{-1}$ Permutations of their rows give additional morphisms. Can they be expressed as tensor product of $|1\rangle$ and $|2\rangle$? The SVD of G_{12} gives a decomposition as a linear combination of tensor products, not necessarily involving any of the $|n\rangle$ -they may even have negative components, i.e., not a partition of N at all.

C. Examples of partitions as $SO(N)$ rotations

To illustrate these result, let's consider all possible partitions of 3 elements. For the partition $\bar{0} = 1|2|3$ we have the trivial result that the number operator is the identity. While any rotation $\mathbf{U}$ satisfies Eq.(62), its precise definition (59) requires it to be $\mathbf{U} = \mathbf{1}_{N\times N}$. Hence, the space of all possible partitions of N elements includes the identity rotation. For the partition $P_2 = 1|23$, the pseudo-ladder and number operators are

$$\mathbf{a}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \hat{\mathbf{N}}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{73a}$$

the former having a left null-space spanned by $(0, -1, 1)$. It follows that the adjacency matrix associated to the partition $\mathbf{A}(P_2)$ corresponds to a 45° rotation of the cluster-size distribution given by $\hat{\mathbf{N}}_2$ around cluster $\overrightarrow{a}^\dagger_{i=1} = (1, 0, 0)$,

$$\mathbf{U}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \tag{73b}$$

$$\mathbf{A}(P_2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}. \tag{73c}$$

The reduced information contained in $\hat{\mathbf{N}}$, namely only the cluster sizes, as opposed to their detailed memberships, introduces the ambiguity that the other partitions of two clusters, $P_2' = 2|13$ and $P_2'' = 3|12$, have exactly the same size distribution $\hat{\mathbf{N}}_2$ -modulo trivial permutations of its columns . The corresponding rotations

$$\mathbf{U'}_2 = \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \quad \mathbf{U''}_2 = \begin{bmatrix} 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{bmatrix} \tag{74a}$$

yield the correct adjacency matrices

$$\mathbf{A}(P_2') = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \mathbf{A}(P_2'') = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{74b}$$

Note that the rotations

$$\mathbf{U'}_{2'} = \begin{bmatrix} 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \quad \mathbf{U''}_{2'} = \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 1 & 0 & 0 \end{bmatrix} \tag{74c}$$

also yield the expected adjacency matrices, but their determinants are -1. Finally, partition $\bar{1} = 123$ corresponds to

$$\hat{\mathbf{N}}_1 = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \mathbf{U}_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \end{bmatrix} \quad \mathbf{A}(\bar{1}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \tag{75a}$$

The definition of the rotation $\mathbf{U}$ requires the same sorting of the cluster (column) vectors as that for $\hat{\mathbf{N}}$: If instead we perform a rotation on $\hat{\mathbf{N}}_2$ similar to $\mathbf{U}_2$ but that doesn't respect that order, e.g,

$$\mathbf{U'}_\pm = \begin{bmatrix} \mp 1/\sqrt{2} & 0 & \pm 1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \quad \mathbf{U''}_0 = \begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix}, \tag{76a}$$

the corresponding *adjacency matrix* would be

$$\mathbf{A'}_\pm = \begin{bmatrix} 1/2 & 0 & \mp 1/2 \\ 0 & 2 & 0 \\ \mp 1/2 & 0 & 1/2 \end{bmatrix}, \tag{76b}$$

modulo permutations of its columns and rows. Rotating instead around the third *axes*,

$$\mathbf{U''} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{76c}$$

yields

$$\mathbf{A}'' = \begin{bmatrix} 3/2 & -1/2 & 0 \\ -1/2 & 3/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{76d}$$

We see that none of the matrices $\mathbf{A'}_{\pm}$ nor $\mathbf{A}''$ can be interpreted as the adjacency matrix of a real (crisp) partition according to definition Eq.(15). This is because none of their corresponding rotations $U'_{\pm}$, U''_0 nor U'' respect the column order of $\hat{\mathbf{N}}$.

As a final examples, we will consider the partition of 5 elements $P_5 = 123|45$, where $\mathbf{N} = diag(3,2,0,0,0)$. In this case, the null space of $\mathbf{a}^\dagger$ is 3-dimensional spanned by

$$\begin{aligned} \hat{b}_1 &= (\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}, 0, 0) \\ \hat{b}_2 &= (\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0) \\ \hat{b}_3 &= (0, 0, 0, \frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \end{aligned} \tag{77}$$

while the orthonormalized clusters are given by

$$\begin{aligned} \hat{a}_1 &= (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, 0, 0) \\ \hat{a}_2 &= (0, 0, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}). \end{aligned} \tag{78}$$

Both rotations

$$\mathbf{U}_5 = [\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2, \hat{b}_3] \tag{79}$$

$$\mathbf{U'}_5 = [\hat{a}_1, \hat{a}_2, \hat{b}_3, \hat{b}_1, \hat{b}_2] \tag{80}$$

are sound $SO(3)$ rotations representing partition P_5.

(Mon Apr 8, 2019)

Q 6. *Rotations in $\mathbb{R}^3$ are given by $\mathbb{RP}^3$, i.e., the quaternions[1] $\boldsymbol{q} \in \mathbb{H}$ in $\mathbb{S}^3$ modulo anti-podal points, with $SU(2)$ a double-cover of it. Our $\boldsymbol{U}$ is not such a quaternion. What are the quaternions definition of a partition rotation? Does this shed any new light on what rotations they are? See Stillwell, The Four pillars of Geometry*

Q 7. *What rotations do we get for N=2? Are they rotation in $\mathbb{R}^2$, i.e, $SO(2)$!? Presumably the two rations we get are the identity and $\boldsymbol{U}_{\bar{1}2} \equiv \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$, i.e., $\pi/4$ rotation. Can we give a description of the finite (set) of valid rotations? What geometric object is this set? Is it a group?*

[0] $\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\boldsymbol{i} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$, $\boldsymbol{j} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $\boldsymbol{k} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$. For a conversion from rotation matrix to quaternion, see wikipedia. $q_0 = \frac{1}{2}\sqrt{1 + Tr(R)}$, $q_i = \frac{1}{4q_0}(-\epsilon_{ijk})\, R_{[j,k]}$.

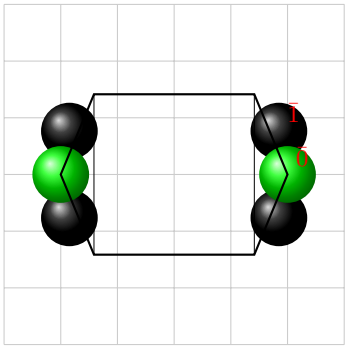


FIG. 4: All actual partition rotations plus the quaternion $\boldsymbol{j}$ do form a group, but together they constitute a valid set of rotations only for partition $\bar{1}$. That seems true unless we drop the interpretation of $\boldsymbol{U}$ as list of *normalized column-clusters*. The fundamental relation (62) doesn't actually require it. With this new freedom, the Hasse space of $\Pi(2)$ seems to map to the (projective) space of vertices of this octagon.

What if we count all matrices $\boldsymbol{U}$ giving rise to a valid adjacency matrix, i.e. $\forall \boldsymbol{U} \in O(2)$, what is the geometry of valid "rotations"? ± 1, $\frac{\pm}{2}(\sqrt{2-\sqrt{2}} \pm \mathbf{j}\sqrt{2+\sqrt{2}})$?

*By partitions, however, $\bar{1} = 12$ has the geometric interpretation of **2 points** $-1, +1$, while $\bar{0} = 1|2$, a **square**, both with a **twofold multiciplicity**. This way of assigning a geometric interpretations gives a **double-cover per partition**. Uhm...*

*Alas, **the set of partitions as rotations (e.g. these quaternions for N=2) do not form a group!** We are missing the quaternion $\boldsymbol{j}$. This quaternion is a valid rotation for $\bar{0}$, but not for $\bar{1}$ as $\boldsymbol{j} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ and*

$$\boldsymbol{j}\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\boldsymbol{j}^\dagger = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{81}$$

which is not a valid adjacency matrix for that partition. That is, we recover a group of valid rotations only for the trivial case of partition $\bar{0}$. Hence, it doesn't seem useful to pursue this analogy further.

That seems true unless we drop the interpretation of $\boldsymbol{U}$ as list of normalized column-clusters. The fundamental relation (62) doesn't actually require it. With this new freedom, the Hasse space of $\Pi(2)$ seems to map to the (projective) space of vertices of this octagon.

Q 8. *What is the geometry of the bipartioning of 3 elements? From the case of $N = 2$ this geometric interpretation of partitions seems to be a subset of $O(2, \mathbb{F}_q) \subset \mathbb{S}^1$. For $N = 2$, $q = 2$.*

Q 9. *The octagon is a multiplicative group, and a finite, 2-dim vector space by addition. Can we related to our idea of a "partition-as-a-ket"?*

Q 10. *We always played with the posiblity to use a density matrix description. A problem with that was to make sense of a linear combination of states. Does the octagon*

represent those states and their linear combinations? The fact that several states represent the same adjacency matrix $\boldsymbol{A}$ sounds like $\boldsymbol{A}$ is "the observable" to "keep track of"; the states, $\boldsymbol{U}$, can change ($|\pm\rangle = (\mathbb{1} \pm \boldsymbol{j})$). Furthermore, linear combinations or products (as a group) give rise to maps between different Hasse levels.

(Mon Apr 15 2019) Caution!: If we fix the order of the clusters in $\hat{\boldsymbol{N}}$, **then there is ONLy 1 such rotation $\boldsymbol{U}$**.

$\bar{0} = 1|2$: The geometric object is isomorphic with $\mathbb{S}^1$, i.e., all rotations in 2-d fit the bill.

$\bar{1} = 12$: The quaternion corresponding to $\boldsymbol{U}_{12}$ is $\boldsymbol{q} = \frac{1}{2}\{\sqrt{2+\sqrt{2}}\,\mathbb{1} + \boldsymbol{k}\sqrt{2-\sqrt{2}}\}$. Here $\theta/2 = 22.5°$, where $\boldsymbol{q} = \cos\theta/2\,\mathbb{1} + \boldsymbol{k}\sin\theta/2$.

The arbitrary choice of rotation axis (i.e., either $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$) doesn't represent any additional degree of freedom.

Changing $\boldsymbol{k} \to -\boldsymbol{k}$, leads to a different $\boldsymbol{U}$, which entails only a permutation of the cluster columns. Hence, we are also not free to change once fixed the order of columns/rows in $\hat{\boldsymbol{N}}$.

Instead of finding the equivalent quaternion corresponding to rotation $\boldsymbol{U}_{12}$, we could directly interpret this rotation as a quaternion $\boldsymbol{q} = \frac{1}{\sqrt{2}}\{\mathbb{1} - \boldsymbol{j}\}$. However, the interpretation as a rotation is that of an angle of 90° around the $\boldsymbol{j}$-axis. uhm...It then seems arbitrary the appearance of the quaternion $\boldsymbol{j}$ instead any of the other two...

In summary: NO OCTAGON!

1|23: $\boldsymbol{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$. We have

$$q_0 = \frac{1}{2}\{1+1+\sqrt{2}\}^{1/2} = \frac{1}{2}\sqrt{2+\sqrt{2}}$$
$$q_2 = q_3 = 0$$
$$q_1 = \frac{1}{4q_0}(-\epsilon_{123})\boldsymbol{U}_{[2,3]} = \frac{1}{2\sqrt{2+\sqrt{2}}}(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}) = \frac{\sqrt{2-\sqrt{2}}}{2}$$

Hence, $\boldsymbol{q}_{1|23} = \frac{1}{2}\{\sqrt{2+\sqrt{2}}\,\mathbb{1} + \boldsymbol{i}\sqrt{2-\sqrt{2}}\}$.

The reflection, $\boldsymbol{i} \to -\boldsymbol{i}$ leads again to permutation of the cluster columns:

$$q \equiv_{1|23} (\boldsymbol{i} \to -\boldsymbol{i}) = \frac{1}{2}\{\sqrt{2+\sqrt{2}}\,\mathbb{1} - \boldsymbol{i}\sqrt{2-\sqrt{2}}\}$$
$$q^\dagger = \frac{1}{2}\{\sqrt{2+\sqrt{2}}\,\mathbb{1} + \boldsymbol{i}\sqrt{2-\sqrt{2}}\}$$
$$\boldsymbol{q}\boldsymbol{i}\boldsymbol{q}^\dagger = \frac{\boldsymbol{q}}{2}\{\sqrt{2+\sqrt{2}}\,\boldsymbol{i} - \sqrt{2-\sqrt{2}}\,\mathbb{1}\} = \boldsymbol{i}$$
$$\boldsymbol{q}\boldsymbol{j}\boldsymbol{q}^\dagger = \frac{\boldsymbol{q}}{2}\{\sqrt{2+\sqrt{2}}\,\boldsymbol{j} - \sqrt{2-\sqrt{2}}\,\boldsymbol{k}\} = \frac{1}{2}\{\boldsymbol{j} - \boldsymbol{k}\}$$
$$\boldsymbol{q}\boldsymbol{j}\boldsymbol{q}^\dagger = \frac{\boldsymbol{q}}{2}\{\sqrt{2+\sqrt{2}}\,\boldsymbol{k} + \sqrt{2-\sqrt{2}}\,\boldsymbol{j}\} = \frac{1}{2}\{\boldsymbol{j} + \boldsymbol{k}\}$$
$$\boldsymbol{U} \to \boldsymbol{U}' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

2|13: $\boldsymbol{U}_{2|13} = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$. We have

$$q_0 = \frac{1}{2}\{1 - \frac{1}{\sqrt{2}}\}^{1/2} = \frac{1}{2\sqrt{2}}\sqrt{2-\sqrt{2}}$$
$$q_1 = \frac{1}{4q_0}(-\epsilon_{123})\boldsymbol{U}_{[2,3]} =$$
$$q_2 = \frac{1}{4q_0}(-\epsilon_{123})\boldsymbol{U}_{[2,3]} =$$
$$q_3 = \frac{1}{4q_0}(-\epsilon_{123})\boldsymbol{U}_{[2,3]} = \frac{1}{2\sqrt{2+\sqrt{2}}}(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}) = \frac{\sqrt{2-\sqrt{2}}}{2}$$

Hence, $\boldsymbol{q}_{2|13} = \frac{1}{2\sqrt{2}}\{\sqrt{2-\sqrt{2}}\,\mathbb{1} + \boldsymbol{i}\sqrt{2+\sqrt{2}} + \boldsymbol{j}\sqrt{2+\sqrt{2}} + \boldsymbol{k}\sqrt{2-\sqrt{2}}\}$.

This corresponds to $\cos\theta/2 = \frac{\sqrt{2-\sqrt{2}}}{2\sqrt{2}} \to \theta \approx 98.42°$ and $\hat{n} = \frac{1}{\sqrt{6-\sqrt{2}}}\{\sqrt{2+\sqrt{2}}\,\boldsymbol{i} + ...\}$ and $\widehat{\hat{n}\boldsymbol{i}} \sim 30.36° = \widehat{\hat{n}\boldsymbol{j}}$ $\widehat{\hat{n}\boldsymbol{k}} \sim 69.06°$

D. Tensorial form of the fundamental relation

The previous examples show what seems are general results: while $\hat{\mathbf{N}}$ in Eq.(62) considers S as a set of *indistinguishable elements* (*particles*), the correct label assignment to each element is provided by the information contained in the rotation $\mathbf{U}$. The sorting of the columns in $\mathbf{U}$ that represent clusters has to be consistent with that in $\hat{\mathbf{N}}$. Finally, we see that to any given pair of equivalence relation $\mathbf{A}$ and size distribution $\hat{\mathbf{N}}$ there corresponds two possible family of rotations $\mathbf{U}_\pm \in O(N)$ mapping one into the other, namely one with determinant $\det(\mathbf{U}_+) = +1$ and the other having $\det(\mathbf{U}_-) = -1$. These correspond to the remaining degree of freedom in choosing the orientation of the basis $\{\hat{b}_i\}$ of the null-space of $\mathbf{a}^\dagger$. As it is only the connected part of $O(N)$, $SO(N)$,

the one which contains the identity, by definition we will consider rotations $\mathbf{U} \in SO(N)$ only, that is, those with unit determinant. This representation $\mathbf{U}$ will then be unique modulo any rotation that preserves the orientation of the null space of $\mathbf{a}^\dagger$.

Using the tensor product of vectors we can express definition (15) directly in terms of the cluster vectors as

$$\mathbf{A} = \sum_i \overrightarrow{a}_i \otimes \overrightarrow{a}_i{}^\dagger = \sum_i n_i \, \hat{a}_i \otimes \hat{a}_i^\dagger \tag{82a}$$

where the second equality is exactly Eq.(62). The isometry matrix is given by $\mathbf{U} = \left[\cdots \hat{a}_i \cdots \hat{b}_j \cdots\right] \equiv \left[\cdots \hat{u}_i \cdots\right]$. A more general expression is

$$\mathbf{A} = \sum_{ij} g_{ij} \, \hat{u}_i \otimes \hat{u}_j^\dagger, \tag{82b}$$

where it is assumed that $g_{ij} = g_{ji} = 0$ if $i > \mathcal{K}$.

For the case of a crisp partition with cluster-size distribution $\hat{\mathbf{N}} = diag(\{n_i\})$ and writing its general symmetry rotation as $\mathbf{\Lambda} = [\hat{\Lambda}_1 \cdots \hat{\Lambda}_N]$, with $\{\hat{\Lambda}_i\}$ an orthonormal set of vectors, the occupation number matrix is

$$\hat{\mathbf{N}} = \sum_i n_i \, \hat{\Lambda}_i \otimes \hat{\Lambda}_i^\dagger. \tag{83}$$

E. Partitions on a graph

1. *Cluster decomposition of $\boldsymbol{\omega}$*

Usually, a clustering is defined on a graph $(S, \boldsymbol{\omega})$ of weights $\boldsymbol{\omega}_{\alpha\beta}$ between nodes of a set S. We will assume that the graph has no self-loops, i.e., that $\boldsymbol{\omega}_{\alpha\alpha} \equiv 0$. In this section we will introduce some definitions that may become useful later.

The adjacency matrix $\boldsymbol{\mathcal{A}}$ of the graph is equivalent to $\boldsymbol{\omega}$ by setting all *relevant* weights to 1 and the rest to 0, and represents just the topological content of the $\boldsymbol{\omega}$. Those weights that are relevant will depend on the given graph and may be defined through a *cut-off* condition as

$$\Xi(\boldsymbol{\omega}_{\alpha\beta}) = 0. \tag{84a}$$

Then $\boldsymbol{\mathcal{A}}(\boldsymbol{\omega})$ is defined as

$$\boldsymbol{\mathcal{A}}(\boldsymbol{\omega})_{\alpha\beta} \equiv 1 - \delta_{0\,\Xi(\boldsymbol{\omega}_{\alpha\beta})}. \tag{84b}$$

For the sake of simplicity, we may think of the simplest case given by

$$\Xi(\boldsymbol{\omega}_{\alpha\beta}) = \boldsymbol{\omega}_{\alpha\beta} \tag{85a}$$
$$\boldsymbol{\mathcal{A}}(\boldsymbol{\omega})_{\alpha\beta} = 1 - \delta_{0\,\boldsymbol{\omega}_{\alpha\beta}}. \tag{85b}$$

As mentioned above, $\mathbf{a}^\dagger$ provides the mapping between the space of elements and that of clusters. Let's work out this projection more in detail. Given a partition P on a graph $\boldsymbol{\omega}$, we define the *intra-cluster* weights as

$$\boldsymbol{\omega}^<_{\alpha\beta} \equiv \boldsymbol{\omega}_{\alpha\beta} \, \delta_{P(\alpha)P(\beta)} = \boldsymbol{\omega}_{\alpha\beta} \, \overrightarrow{a}_\alpha \cdot \overrightarrow{a}^\dagger_\beta, \tag{86a}$$

where $\overrightarrow{a}_\alpha$ is the α-th row vector of matrix $\mathbf{a}$. Similarly, we define the *inter-cluster* weights as

$$\boldsymbol{\omega}^>_{\alpha\beta} \equiv \boldsymbol{\omega}_{\alpha\beta} \left[1 - \overrightarrow{a}_\alpha \cdot \overrightarrow{a}^\dagger_\beta\right], \tag{86b}$$

such that any weight can be decomposed uniquely as $\boldsymbol{\omega}_{\alpha\beta} = \boldsymbol{\omega}^<_{\alpha\beta} + \boldsymbol{\omega}^>_{\alpha\beta}$.

The sum of all weights within cluster i, W_i, is given by

$$W_i = \frac{1}{2} \overrightarrow{a}^\dagger_i \boldsymbol{\omega} \overrightarrow{a}_i, \tag{87a}$$

where $\overrightarrow{a}_i$ is the i-th column vector of matrix $\mathbf{a}$ and the factor $\frac{1}{2}$ corrects the double-counting of edges between any pair of elements α and β. The sum of all intra-cluster weights can be easily expressed in matrix notation as

$$\sum_i W_i = \frac{1}{2} Tr\left(\mathbf{a}^\dagger \boldsymbol{\omega} \mathbf{a}\right). \tag{87b}$$

In the same way, we can write the sum of all weights between clusters i and j as

$$\boldsymbol{W}_{ij} \equiv \overrightarrow{a}^\dagger_i \boldsymbol{\omega} \overrightarrow{a}_j. \tag{87c}$$

In matrix notation it is $\boldsymbol{W} \equiv \mathbf{a}^\dagger \boldsymbol{\omega} \mathbf{a}$, which in tensorial form reads

$$\boldsymbol{W} = \sum_{\alpha\beta} \boldsymbol{\omega}_{\alpha\beta} \, \overrightarrow{a}_\alpha \otimes \overrightarrow{a}^\dagger_\beta. \tag{87d}$$

Given a graph $(S, \boldsymbol{\omega})$, the total weight

$$\omega \equiv \sum_{\alpha\beta} \boldsymbol{\omega}_{\alpha\beta} = \sum_{\alpha\beta} \left(\boldsymbol{\omega}^<_{\alpha\beta} + \boldsymbol{\omega}^>_{\alpha\beta}\right) \tag{88}$$

is invariant under any change of partition P of S. ω can be decomposed in terms of clusters as

$$\begin{aligned} \omega = W &\equiv \frac{1}{2} \sum_{ij} \boldsymbol{W}_{ij} \\ &= \sum_i W_i + \sum_{i<j} \boldsymbol{W}_{ij} \\ &\equiv \sum_i \phi_i, \end{aligned} \tag{89}$$

where $\phi_i \equiv W_i + \frac{1}{2} \sum_{i \neq j} \boldsymbol{W}_{ij}$.

2. *Intrinsic spin variables: Newman's modularity*

In the search for a variational approach to the clustering problem, a common approach has been to look for an ad-hoc order-disorder like hamiltonian (See [17, 58, 59] and references therein). This raises the following question:

Q 11. *Are there some intrinsic* spin *variables one may identify in a graph* $(S, \boldsymbol{\omega})$ *upon which we can define such a variational approach?*

The prior information on the problem is in principle very simple, namely that contained in the graph $\boldsymbol{\omega}$. Let's thus try to identify what could be some natural, intrinsic spin/momenta variables based solely on that information.

To each node α we can assign a N-dimensional vector $\overrightarrow{\mu}_\alpha$, which we call the *topological moment* of node α, given by all edges spanned by α. , i.e.,

$$(\overrightarrow{\mu}_\alpha)^\gamma \equiv \boldsymbol{\mathcal{A}}(\omega)^\gamma_\alpha. \tag{90a}$$

It is straightforward to see that $\mu^2_\alpha \equiv \|\overrightarrow{\mu}_\alpha\|^2 = \sum_\beta \boldsymbol{\mathcal{A}}_\alpha{}^\beta \boldsymbol{\mathcal{A}}_{\beta\alpha} = k_\alpha$, where k_α is the degree of node α, and

$$\overrightarrow{\mu} \equiv \sum_\alpha \overrightarrow{\mu}_\alpha = \overrightarrow{k}, \tag{90b}$$

is the total topological moment of the graph. Then we have

$$\overrightarrow{\mu}_\alpha \cdot \overrightarrow{\mu}_\beta = \#\text{ Paths of length 2 from } \alpha \text{ to } \beta, \tag{90c}$$

and in general it is

$$\overrightarrow{\mu}^\dagger_\alpha \boldsymbol{\mathcal{A}}^l \overrightarrow{\mu}_\beta = \#\text{ Paths of length } 2+l \text{ from } \alpha \text{ to } \beta. \tag{90d}$$

A *self-loop* of an element α can be assigned a moment $\overrightarrow{\mu}^o_\alpha$ which coincides with the base vector representing the element itself e_α. Thus, even if the graph $\boldsymbol{\omega}$ has no self-loops, we can view e_α as the moment of a *virtual* self-loop $\overrightarrow{\mu}^0_\alpha$ that is related to the moment of α as $\boldsymbol{\mathcal{A}}(\boldsymbol{\omega})\overrightarrow{\mu}^0_\alpha = \boldsymbol{\mathcal{A}}(\boldsymbol{\omega})e_\alpha = \overrightarrow{\mu}_\alpha$. This allows us to express the number of paths of length l between elements α and β in the usual form namely as powers of the adjacency matrix of the given graph,

$$\overrightarrow{\mu}^0_\alpha \boldsymbol{\mathcal{A}}^l(\boldsymbol{\omega})\overrightarrow{\mu}^0_\beta. \tag{90e}$$

For $l = 1$ and $\alpha = \beta$ this gives the number of actual self-loops for each node α, which we assume to be zero, i.e., $\overrightarrow{\mu}^0_\alpha \boldsymbol{\mathcal{A}}(\boldsymbol{\omega})\overrightarrow{\mu}^0_\alpha = 0$.

Furthermore, $\mathbf{a}^\dagger$ maps $\overrightarrow{\mu}_\alpha$ to a $\mathcal{K}$-dimensional vector $\overrightarrow{S}_\alpha$, which we call the *spin* of node α, where each component i is the number of edges starting on α and ending in cluster i, as given by

$$\overrightarrow{S}_\alpha \equiv \mathbf{a}^\dagger \overrightarrow{\mu}_\alpha = [\overrightarrow{a}_i \cdot \overrightarrow{\mu}_\alpha] \tag{91a}$$

Thus, the mapping $\overrightarrow{a}^\dagger$ represents a change of representation from one based on nodes to one based on clusters. It is straightforward to see that $\overrightarrow{\Sigma}_\alpha \equiv \mathbf{a}\overrightarrow{S}_\alpha$ represents the same information in the space of all nodes, i.e., each component β,

$$\left(\overrightarrow{\Sigma}_\alpha\right)^\beta = \left(\mathbf{a}\overrightarrow{S}_\alpha\right)^\beta = \delta^{P(\beta)}_{P(\gamma)} \boldsymbol{\mathcal{A}}(\omega)^\gamma_\alpha, \tag{91b}$$

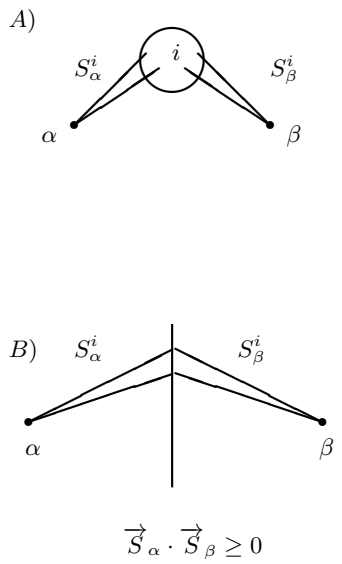


FIG. 5: Graphical representation of the intrinsic spin variables in the graph clustering problem. *A)* S^i_α (S^i_β) is total number of edges between node α (β) and cluster i. *B)* The scalar product of two spins yields the sum of all 2-edges paths joining nodes α and β. The vertical line denotes the summation over all clusters.

is the number of edges starting at node α and ending at the cluster containing node β.

Newman and Girvan's modularity measure is given by[19]

$$Q \equiv \sum_i \left[\frac{e_i}{M} - \left(\frac{\sum_{\alpha\in C_i} k_\alpha}{2M}\right)^2\right], \tag{92a}$$

where e_i is the number of intra-cluster edges for cluster C_i and $M = \sum_{\alpha\leq\beta} \boldsymbol{\mathcal{A}}_{\alpha\beta}$ is the total number of edges in the graph. Hence, the different terms in (92a) can be rewritten as

$$\sum_{\alpha\in C_i} k^\alpha = \overrightarrow{a}_i \cdot \overrightarrow{k} = \sum_\alpha S^i_\alpha \tag{92b}$$

$$\sum_i \left(\sum_{\alpha\in C_i} k^\alpha\right)^2 = \sum_i \left(\overrightarrow{a}_i \cdot \overrightarrow{k}\right)^2 = \sum_{\alpha\,\beta} \overrightarrow{S}_\alpha \cdot \overrightarrow{S}_\beta = \\ = \left(\mathbf{a}^\dagger \overrightarrow{\mu}\right) \cdot \left(\mathbf{a}^\dagger \overrightarrow{\mu}\right) \equiv \overrightarrow{S} \cdot \overrightarrow{S}, \tag{92c}$$

where $\overrightarrow{S}$ is the total spin of the graph, and the modularity is

$$Q = -\mathcal{H}(S) \equiv \frac{1}{2M}\sum_\alpha \overrightarrow{S}_\alpha \cdot \overrightarrow{a}_\alpha - \frac{1}{4M^2}\sum_{\alpha\,\beta} \overrightarrow{S}_\alpha \cdot \overrightarrow{S}_\beta \tag{93}$$

which has a structure akin to that of a spin-system Hamiltonian with long-range interactions in the presence

of an external field $\overrightarrow{a}_\alpha$. A similar conclusion was already stated in [17] although they started from an ad-hoc Potts Hamiltonian that already conveyed the form of the modularity. It is straightforward to see that the second term in Eq.(93) is always negative, vanishing only for the trivial case of a network $\mathcal{A}(\boldsymbol{\omega})$ with no edges. Technically, the optimal partition would then remain undertermined. We will then convey to choose partition $\bar{0}$ as the optimal partition. For any two elements α and β, this term is more negative the more edges these elements have that end in the same cluster. It, therefore, tends to split the system into as many clusters as possible. Conversely, the first term attains its maximum value when each node has edges only within the same cluster it belongs to, i.e., it favors partition $\bar{1}$, the clustering of all elements together. This spin model thus shares some similirities with that of an antiferromagnet, but as opposed to usual treatments of the latter model, the spin-spin interactions in Eq.(93) contain all pairs of spins. One is tempted to consider it then as a model with long-range interactions. This, however, seems somewhat misleading, as it is not clear what the meaning of distance on the network is, nor what the dimensionality of the system is.

As S^i_α depends itself on $\mathbf{a}$, the product $\overrightarrow{S}_\alpha \cdot \overrightarrow{a}_\alpha$ does not correspond exactly to what is usually considered as the action of an external field on a local spin variable. However, we may uncouple spin and external field by writing the modularity in terms of the topological momenta.

Let us consider the action of a partition rotation $\mathbf{U}$ on the topological momentum $\overrightarrow{\mu}_\alpha$ of an element α.

$$\begin{aligned} \mathbf{U}^\dagger \overrightarrow{\mu}_\alpha &= \hat{\mathfrak{a}}^\dagger \overrightarrow{\mu}_\alpha + \hat{\mathfrak{b}}^\dagger \overrightarrow{\mu}_\alpha \\ &\equiv \overrightarrow{\sigma}_\alpha + \overrightarrow{\tau}_\alpha, \end{aligned} \tag{94}$$

which defines the *dispersion spin* $\overrightarrow{\sigma}_\alpha = \hat{\mathfrak{a}}^\dagger \overrightarrow{\mu}_\alpha$ and *hole spin* $\overrightarrow{\tau}_\alpha = \hat{\mathfrak{b}}^\dagger \overrightarrow{\mu}_\alpha$. See figure (6). It is straightforward to see that

$$\overrightarrow{\sigma}^\dagger_\alpha \cdot \overrightarrow{\sigma}_\beta = \sum_i \overrightarrow{\mu}^\dagger_\alpha \hat{a}_i \otimes \hat{a}^\dagger_i \overrightarrow{\mu}_\beta = \sum_i \frac{S^i_\alpha S^i_\beta}{n_i}, \tag{95}$$

where its module is bound by $|\sigma^i_\alpha| \lesssim \sqrt{n_i}$ and $\|\overrightarrow{\sigma}_\alpha\| \lesssim \mathcal{K}\sqrt{n_i}$. We have then

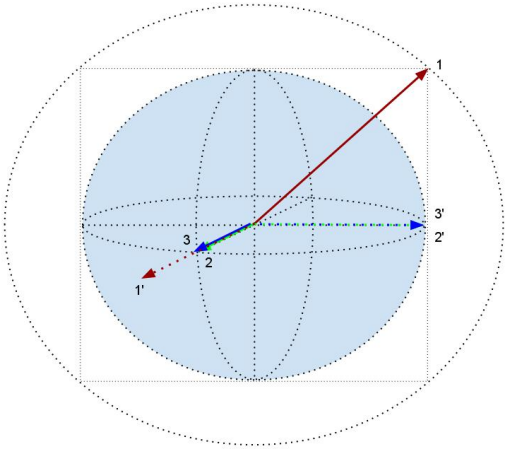
$$\overrightarrow{\sigma}^\dagger_\alpha \hat{\mathbf{N}} \overrightarrow{\sigma}_\beta = \overrightarrow{\mu}^\dagger_\alpha \mathbf{A} \overrightarrow{\mu}_\beta = \sum_i S^i_\alpha S^i_\beta. \tag{96}$$

The following projection yields the action of $\overrightarrow{a}_\alpha$ on the spin of element β

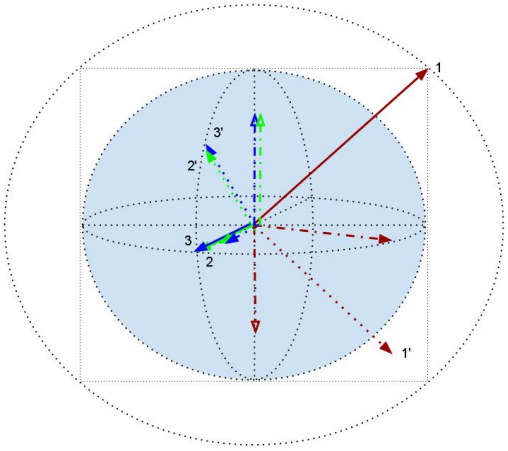
$$\overrightarrow{\sigma}^{o\dagger}_\alpha \hat{\mathbf{N}} \overrightarrow{\sigma}_\beta = \overrightarrow{\mu}^0_\alpha \mathbf{A} \overrightarrow{\mu}_\beta = \sum_i (\overrightarrow{a}_\alpha)^i S^i_\beta. \tag{97}$$

Then we can write the modularity as

$$Q = -\mathcal{H} \equiv \frac{1}{2M} \sum_\alpha \overrightarrow{\mu}_\alpha \cdot \overrightarrow{a}_{P(\alpha)} - \frac{1}{4M^2} \sum_{\alpha\,\beta} \overrightarrow{\mu}_\alpha \mathbf{A} \overrightarrow{\mu}_\beta, \tag{98}$$



(a) P=23|1



(b) P=12|3

FIG. 6: Spin rotations. The effect of two different partition rotations on the topological momenta (continuous arrows) of 3 elements is shown for two partitions defined on the same network $2-1-3$: (6a) $P = 23|1$ and (6b) $P = 12|3$. The rotated momentum $\mathbf{U}^\dagger \overrightarrow{\mu}_\alpha$ is shown in dotted lines. Dash-dotted arrows indicate the spin $\overrightarrow{\sigma}_\alpha$ (filled arrowheads) and the hole spin (hollow arrowhead) $\overrightarrow{\tau}_\alpha$, such that $\overrightarrow{\mu}'_\alpha = \mathbf{U}^\dagger \mu_\alpha = \overrightarrow{\sigma}_\alpha + \overrightarrow{\tau}_\alpha$, where we define $\overrightarrow{\sigma}_\alpha \equiv \hat{\mathfrak{a}}^\dagger \mu_\alpha$ and $\overrightarrow{\tau}_\alpha \equiv \hat{\mathfrak{b}}^\dagger \mu_\alpha$. Because of symmetry, the spins of elements 2 and 3 coincide.

where now the external field $\overrightarrow{a}_{P(\alpha)}$ is homogenous over the elements belonging to the same cluster. This can be written in a more compact form as

$$Q = -\mathcal{H} = r \sum_\alpha \overrightarrow{\mu}_\alpha *_{\mathbf{A}} \overrightarrow{\mu}^o_\alpha - r^2 \sum_{\alpha\beta} \overrightarrow{\mu}_\alpha *_{\mathbf{A}} \overrightarrow{\mu}_\beta, \tag{99}$$

with $r = \frac{1}{2M}$ and $\overrightarrow{v} *_{\mathbf{A}} \overrightarrow{w} \equiv \langle \overrightarrow{v} | \overrightarrow{w} \rangle_{\mathbf{A}} = \mathbf{A}_{\alpha\beta} v^\alpha w^\beta$, where $\mathbf{A}(P)$ plays the role of a pseudo-metric in the vectorial space of momenta of the nodes of graph $\boldsymbol{\omega}$ (see Eq.(28)).

Let us consider two partition $\mathbf{A}^n$ and $\mathbf{A}^m$ corresponding to the same cluster-size distribution $\hat{\mathbf{N}}$. As we have seen above, there is a relative rotation $\mathbf{R}_{mn}$ such that $\mathbf{A}^m = \mathbf{R}_{mn}\mathbf{A}^n\mathbf{R}^\dagger_{mn}$. Therefore, the modularity Q^m corresponding to partition $\mathbf{A}^m$ can be seen as the modularity of a partition $\mathbf{A}^n$ with a network whose momenta are rotated by $\mathbf{R}^\dagger_{mn}\overrightarrow{\mu}_\alpha$ relative to the original network. This suggests the possibility of an alternative way of looking at the partitioning problem of a complex network: instead of searching for the optimal partition given a fixed topology, we can view it as searching for the optimal topology for a fixed partition, albeit the network now becomes a weighted network. For such a view to hold, it would be necessary to find a positive answer to the following question:

Q 12. *Given two partitions corresponding to two different cluster-size distributions, can we find a rotation that transforms them into each other?*

As shown in question ***Q***5, this has a partially positive answer. Indeed, the similarity transformation between partitions with different cluster-size distributions is given by an isometry of $\boldsymbol{l_1}$ and not $\boldsymbol{l_2}$.

The present result (98) has been obtained without any ad-hoc assumptions of the number of spin states nor other parameters, contrary to the departing Ansatz of [17], and it simply shows that the modularity of Newman and Girvan is in essence but an order-disorder like Hamiltonian. Moreover, (93) identifies what is a genuine definition of the spin of a node $\overrightarrow{S}_\alpha$ relevant to the clustering problem, namely as a projection onto the cluster space of a column of the adjacency matrix of the graph $\boldsymbol{\omega}$. We will be able to substantiate such a claim only if we can show that the Hamiltonian $\mathcal{H}(S)$ can effectively be used to derive a sound clustering algorithm, much like it was done in [17].

In the quest for finding a sound (spin) Hamiltonian analog for the clustering problem, Reichardt and Bornholdt...

V. SYMMETRY OF COMMUNITIES IN COMPLEX NETWORKS

We will demonstrate in this section how the concept of ideal (boolean) partition bears in essence the recognition of a symmetry in the distribution of weights over the set of clusters given by $\boldsymbol{W}$.

A. The partition Lie algebra $\mathfrak{p}(N,\mathcal{K})$

As per Eq.(87d), this distribution has a tensorial representation given by $\boldsymbol{W} = \sum_{\alpha\beta}\boldsymbol{\omega}_{\alpha\beta}\eta_{\alpha\beta}$ with

$$\eta_{\alpha\beta} \equiv \overrightarrow{a}_\alpha \otimes \overrightarrow{a}^\dagger_\beta. \tag{100}$$

We will show now that the operators $\eta_{\alpha\beta}$ define a Lie algebra describing the symmetries of the distribution of weights $\boldsymbol{W}$ among clusters $\mathbf{a}$, which we will call the partition Lie algebra. To each Hasse level, i.e., to each number of clusters $\mathcal{K}$ corresponds different representations of this algebra. We will therefore denote this algebra in general by $\mathfrak{p}(N,\mathcal{K})$.

Bilinear products of fermion or boson creation and annihilation operators allow for the introduction of Lie algebras that have been proven to be extremely useful in quatum field theory and many-particle systems [46]. For instance, the isospin can be obtained by considering all bilinear products of neutron and proton creation and annihilation operators which do not change the number of particles. We will show that the generators (100) can also be interpreted in a similar way.

Clearly, the set of operators in Eq.(100) is redundant. For all elements α', β' such that $P(\alpha') = P(\alpha), P(\beta') = P(\beta)$ it is $\eta_{\alpha'\beta'} = \eta_{\alpha\beta}$, as the definition Eq.(100) depends only on the clusters elements α and β belong to. Therefore, it is convenient to change to a notation dependent only on the clusters labels given by

$$\boldsymbol{W} = \sum_{ij} \boldsymbol{W}_{ij}\,\eta_{ij} \tag{101a}$$

$$\boldsymbol{W}_{ij} = \sum_{\alpha\beta} \boldsymbol{\omega}_{\alpha\beta}\delta_{iP(\alpha)}\,\delta_{jP(\beta)} \tag{101b}$$

$$\eta_{ij} \equiv \delta_{iP(\alpha)}\delta_{jP(\beta)}\,\overrightarrow{a}_\alpha \otimes \overrightarrow{a}^\dagger_\beta. \tag{101c}$$

The generators of the partition Lie algebra satisfy the commutation rules

$$[\eta_{\alpha\beta}, \eta_{\alpha'\beta'}] = \lambda^{\mu\nu}_{\alpha\beta\alpha'\beta'}\eta_{\mu\nu} \tag{102a}$$

with structure factors given by

$$\lambda^{\mu\nu}_{\alpha\beta\alpha'\beta'} \equiv \delta^\mu_\alpha\,\delta^\nu_{\beta'}\,\delta_{P(\beta)P(\alpha')} - \delta^\mu_{\alpha'}\,\delta^\nu_\beta\,\delta_{P(\alpha)P(\beta')}. \tag{102b}$$

It is straightforward to prove that the matrices $\eta_{\alpha\beta}$ satisfy the Jacobi identity

$$0 = [\eta_{\alpha\beta}\,,\,[\eta_{\alpha'\beta'}\,,\,\eta_{\alpha''\beta''}]] + [\eta_{\alpha'\beta'}\,,\,[\eta_{\alpha''\beta''}\,,\,\eta_{\alpha\beta}]] + [\eta_{\alpha''\beta''}\,,\,[\eta_{\alpha\beta}\,,\,\eta_{\alpha'\beta'}]]\,. \tag{102c}$$

The following properties hold

$$\eta_{\alpha\beta}\,\eta_{\alpha'\beta'} = \delta_{P(\beta)P(\alpha')}\,\eta_{\alpha\beta'} \tag{103a}$$
$$\eta_{\alpha\beta}^2 = \delta_{P(\alpha)P(\beta)}\,\eta_{\alpha\beta} \tag{103b}$$
$$Tr(\eta_{\alpha\beta}) = \delta_{P(\alpha)P(\beta)} \tag{103c}$$
$$\hat{\boldsymbol{N}} = \sum_{\alpha}\eta_{\alpha\alpha} \tag{103d}$$
$$\left[\hat{\boldsymbol{N}}\,,\,\eta_{\alpha\beta}\right] = (n_\alpha - n_\beta)\,\eta_{\alpha\beta} \tag{103e}$$
$$\left[\hat{\boldsymbol{N}}\,,\,\eta_{\alpha\beta}^2\right] = 0, \tag{103f}$$

with n_α a short-hand notation for the number of elements in the cluster containing α, i.e., $n_\alpha \equiv n_{P(\alpha)}$. We have as well for the structure factors

$$\lambda^{\mu\nu}_{\alpha\beta\alpha'\beta'} \in \{-1\,,\,0\,,\,1\} \tag{104a}$$
$$\lambda^{\mu\nu}_{\alpha\beta\alpha'\beta'} = -\lambda^{\mu\nu}_{\alpha'\beta'\alpha\beta} \tag{104b}$$
$$\lambda^{\mu\nu}_{\alpha\beta\alpha'\beta'}\,\delta_{P(\mu)P(\nu)} = 0 \tag{104c}$$
$$\lambda^{\mu\nu}_{\alpha\alpha\beta\beta} = \left(\delta^\mu_\alpha\delta^\nu_\beta - \delta^\mu_\beta\delta^\nu_\alpha\right)\,\delta_{P(\alpha)P(\beta)}. \tag{104d}$$

The dimension of the matrices $\eta_{\alpha\beta}$ is $\mathcal{K}\times\mathcal{K}$ and there are $\mathcal{K}^2-1$ independent generators, due to condition Eq.(103d), giving rise to the lie algebra $\mathfrak{p}(\mathcal{N},\mathcal{K})\subset\mathfrak{su}(\mathcal{K})$.

The introduction of the partition algrebra $\mathfrak{p}(\mathcal{N},\mathcal{K})$ is a trivial consequence of working on a linear space. The matrix $\mathbf{W}$ belongs in fact to a $\mathcal{K}\times\mathcal{K}$-dimensional linear space, which has a *canonical base* given by the $\mathcal{K}\times\mathcal{K}$ matrices $e_{ij}\equiv\delta_{ij}$. We are free, however, to choose any other set of $\mathcal{K}\times\mathcal{K}$ independent matrices as a base for such a linear space. The generators $\eta_{\alpha\beta}$ simply represent such an example, as do any other combination of them. The following specific examples will help clarifying these points. This trivial nature of the Lie algebra $\mathfrak{p}(\mathcal{N},\mathcal{K})$ means, in particular, that it can be defined in any *vectorial* space with a non-conmutative internal product[55].

The actual relevance of the partition algebra lies in the fact that it allows us to define the ideal partition in terms of a symmetry requirement on the distribution of weights among the different clusters. It is invariance under rotations in the *cluster* space that characterizes the intuitive and widespread idea of *ideal partition*, as we will see later. This concept of symmetry in the clustering problem explains the recurrent approach found hitherto in the literature of using the concept of spin, e.g., in the form of a Pott-model, when looking for suitable mathematical descriptions. Each *spin value* is there tantamount to a cluster label. We will show below that the full concept of spin as an invariant observable under certain transformations can carried over to the partitioning problem.

B. Isospin of the graph-bisectioning problem

Let's consider the lie algebra associated to the case $N=2$ and $\bar{0}_2 = 1|2$. We have the following generators:

$$\eta_{12} = \begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0 & 1\end{pmatrix} = \begin{pmatrix}0 & 1\\0 & 0\end{pmatrix} \tag{105a}$$
$$\eta_{21} = \begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}1 & 0\end{pmatrix} = \begin{pmatrix}0 & 0\\1 & 0\end{pmatrix} \tag{105b}$$
$$\eta_{11} = \begin{pmatrix}1 & 0\\0 & 0\end{pmatrix} \qquad \eta_{22} = \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix} \tag{105c}$$

with the number operator given by $\hat{\boldsymbol{N}} = \eta_{11}+\eta_{22}$. We introduce now the *isospin* operators

$$\boldsymbol{\tau}_+ \equiv \eta_{12} \qquad \boldsymbol{\tau}_- \equiv \eta_{21} \tag{106a}$$
$$\boldsymbol{\tau}_0 \equiv \frac{1}{2}\,(\eta_{11}-\eta_{22})\,. \tag{106b}$$

The properties (103) imply that they all commute with $\hat{\boldsymbol{N}}$ and we also have the commutation rules

$$\begin{aligned}[\boldsymbol{\tau}_0,\boldsymbol{\tau}_+] &= \frac{1}{2}\,[\eta_{11},\eta_{12}] - \frac{1}{2}\,[\eta_{22},\eta_{12}]\\ &= \frac{1}{2}(\eta_{12}+\eta_{12})\\ &= \boldsymbol{\tau}_+\end{aligned} \tag{106c}$$
$$[\boldsymbol{\tau}_0,\boldsymbol{\tau}_-] = -\boldsymbol{\tau}_- \tag{106d}$$
$$[\boldsymbol{\tau}_+,\boldsymbol{\tau}_-] = 2\boldsymbol{\tau}_0. \tag{106e}$$

In quantum mechanics this corresponds to the isospin algebra where $\hat{\boldsymbol{N}}$ is the barion number operator and the charge operator $\boldsymbol{Q}$ is defined through the relation $\boldsymbol{\tau}_0 = \boldsymbol{Q} - \frac{1}{2}\hat{\boldsymbol{N}}$, which in our case is $\boldsymbol{Q} = \eta_{11}$. In the present case, obviously, the *charge of a cluster* is just an arbitrary label for denoting each cluster.

We define the *complex* operators of *angular momentum*

$$\boldsymbol{J}_x \equiv \frac{1}{2}\,(\boldsymbol{\tau}_+ + \boldsymbol{\tau}_-) = \frac{1}{2}\,(\eta_{12}+\eta_{21}) \tag{107a}$$
$$\boldsymbol{J}_y \equiv \frac{1}{2i}\,(\boldsymbol{\tau}_+ - \boldsymbol{\tau}_-) = \frac{-i}{2}\,(\eta_{12}-\eta_{21}) \tag{107b}$$
$$\boldsymbol{J}_z \equiv \boldsymbol{\tau}_0. \tag{107c}$$

It is easy to see that these operators do satisfy the well-known commutation rules of angular momentum

$$[\boldsymbol{J}_x,\boldsymbol{J}_y] = i\boldsymbol{J}_z \tag{107d}$$
$$[\boldsymbol{J}_z,\boldsymbol{J}_x] = i\boldsymbol{J}_y \tag{107e}$$
$$[\boldsymbol{J}_y,\boldsymbol{J}_z] = i\boldsymbol{J}_x, \tag{107f}$$

and that they commute with the casimir operator $\boldsymbol{J}^2 \equiv \boldsymbol{J}_x^2 + \boldsymbol{J}_y^2 + \boldsymbol{J}_z^2$. The latter in addition is related to the (barion) number operator as

$$\boldsymbol{J}^2 = \frac{3}{4}\hat{\boldsymbol{N}} = \frac{1}{2}(\frac{1}{2}+1)\hat{\boldsymbol{N}}. \tag{107g}$$

We see therefore that the partition Lie algebra conveys the expected symmetry which hitherto in the literature had been assigned in an ad-hoc fashion. Indeed, relation (107g) corresponds precisely to that of a system with angular momentum $j = 1/2$, as one would define for any partition with only two clusters. In the Physics literature, Eqs.(106) are known as Schwinger's Oscillator model of Angular Momentum.

Talking about partition $\bar{0}_2$ as having a spin, or angular momentum of $1/2$ is again a short-hand notation for describing the fact that it consists of only two clusters. If this would be the only benefit, introducing these concepts into the clustering problem would certainly be an overkill. The powerful tools of Lie algebras, however, will unveil further consequences of that fact when partioning a complex graph $\boldsymbol{\omega}$. This in turn should provide for a proper justification of their use.

A technical remark is in place here. In rigor, we have not demonstrated that the Lie algebra derived here is that of $\mathfrak{su}(2)$. For that we would need to show that, indeed, the algebra relevant to clustering would be the Lie algebra over the real numbers spanned by $\{i\,\boldsymbol{J}_i\}$. Therefore, so far, we will consider that the Lie algebra unfolded by the bipartitioning problem is (a subalgebra of) $\mathfrak{sl}(2,\mathbb{C})$.

C. $\mathfrak{su}(3)$ symmetry

Let's consider the partition $\bar{0}_3 = 1|2|3$. The partition Lie algebra contains 9 independent operators given by

$$\hat{\boldsymbol{N}} = \eta_{11} + \eta_{22} + \eta_{33} \tag{108a}$$
$$\boldsymbol{\tau}_+ = \eta_{12} \qquad \boldsymbol{\tau}_- = \eta_{21} \tag{108b}$$
$$\boldsymbol{\tau}_0 = \frac{1}{2}\left(\eta_{11} - \eta_{22}\right) \tag{108c}$$
$$\boldsymbol{B}_+ = \eta_{13} \qquad \boldsymbol{B}_- = \eta_{23} \tag{108d}$$
$$\boldsymbol{C}_+ = \eta_{32} \qquad \boldsymbol{C}_- = \eta_{31} \tag{108e}$$
$$\boldsymbol{M} = \frac{1}{3}\left(\eta_{11} + \eta_{22} - 2\eta_{33}\right). \tag{108f}$$

These satisfy the commutation rules of the Lie algebra $\mathfrak{su}(3)$, as we will discuss next. First we note that

$$\begin{aligned}[\boldsymbol{\tau}_0, \boldsymbol{B}_\pm] &= \frac{1}{2}\{[\eta_{11}, \eta_{\pm 3}] - [\eta_{22}, \eta_{\pm 3}]\} \\ &= \frac{1}{2}\eta_{\pm 3} \\ &= \pm\frac{1}{2}\boldsymbol{B}_\pm\end{aligned} \tag{109a}$$
$$[\boldsymbol{\tau}_0, \boldsymbol{C}_\pm] = \pm\frac{1}{2}\boldsymbol{C}_\pm, \tag{109b}$$

where $\eta_{\pm 3}$ is a short-hand notation for η_{13} and η_{23}. Therefore, $\boldsymbol{B}_+$ and $\boldsymbol{C}_+$ *create a (cluster-)*1 *particle and annihilate a (cluster-)2 particle* respectively. These commutation relations are thus consistent with interpreting the tensorial product (100) $\eta_{\alpha\beta} \equiv \overrightarrow{a}_\alpha \otimes \overrightarrow{a}^\dagger_\beta$ as equivalent to the bilinear form $\boldsymbol{a}^\dagger_\alpha \boldsymbol{a}_\beta$ of creation and annihilation operators in Quantum Field Theory. That is, $\eta_{\alpha\beta}$ creates an α *particle/excitation* and *destroys a* β *particle.*

We further have thus that the operators $\boldsymbol{B}_\pm$ annihilate a (cluster-)3 particle and therefore increase $\boldsymbol{M}$ by $+1$, while the operators $\boldsymbol{C}_\pm$ create a (cluster-)3 particle, and thus change $\boldsymbol{M}$ by -1. A simmilar argument leads to the fact that the opertors $\boldsymbol{\tau}_{0,\pm}$, however, leave $\boldsymbol{M}$ unchanged. This yields the additional commutation rules

$$[\boldsymbol{M}, \boldsymbol{B}_\pm] = \boldsymbol{B}_\pm \tag{110a}$$
$$[\boldsymbol{M}, \boldsymbol{C}_\pm] = -\boldsymbol{C}_\pm \tag{110b}$$
$$[\boldsymbol{M}, \boldsymbol{\tau}_{0,\pm}] = 0. \tag{110c}$$

The remaining commutation rules are

$$\begin{aligned}[\boldsymbol{\tau}_\pm, \boldsymbol{B}_\pm] &= [\boldsymbol{\tau}_\pm, \boldsymbol{C}_\pm] \\ &= [\boldsymbol{B}_+, \boldsymbol{B}_-] \\ &= [\boldsymbol{C}_+, \boldsymbol{C}_-] = 0\end{aligned} \tag{110d}$$
$$[\boldsymbol{\tau}_\pm, \boldsymbol{B}_\mp] = \boldsymbol{B}_\pm \tag{110e}$$
$$[\boldsymbol{\tau}_\pm, \boldsymbol{C}_\pm] = -\boldsymbol{C}_\mp \tag{110f}$$
$$[\boldsymbol{B}_\pm, \boldsymbol{C}_\mp] = \frac{1}{2}\left(3\boldsymbol{M} \pm 2\boldsymbol{\tau}_0\right) \tag{110g}$$
$$[\boldsymbol{\tau}_0, \boldsymbol{\tau}_\pm] = \pm\boldsymbol{\tau}_\pm \tag{110h}$$
$$[\boldsymbol{\tau}_+, \boldsymbol{\tau}_-] = \boldsymbol{\tau}_0. \tag{110i}$$

The previous relations define the well-known Lie algebra of $\mathfrak{su}(3)$.

D. Symmetries of the Weight Distribution

Defining a partition on a weighted network $\boldsymbol{\omega}$ entails a particular distribution of these weights among all clusters (communities) given by $\boldsymbol{W}$. It is common in the literature to guide the search for an optimal partition through the following criterion: If $\boldsymbol{\omega}$ represents a measure of similarity, the optimal partition is such that the inter-cluster similarity vanishes (or, eventually, is as low as possible) while the intra-cluster similarity is as high as possible. We shall now show that this criterion can be expressed as a condition of symmetry on $\boldsymbol{W}$.

Let's first consider the case of a bipartitioning, $\mathcal{K} = 2$. From the previous discussion in this section, we see that $\boldsymbol{W}$ can be expressed as a function of the generators of $\mathfrak{sl}(2,\mathbb{C})$, say $\boldsymbol{W} = \boldsymbol{W}(\boldsymbol{J}_i)$. The symmetry properties of

$\boldsymbol{W}$ are given by the following commutators:

$$[\boldsymbol{W}\,,\,\boldsymbol{J}_i] = \sum_{\alpha\,\beta} \omega_{\alpha\,\beta}\,\left[\eta_{\alpha\,\beta}\;\boldsymbol{J}_i\right] \tag{111a}$$

$$\left[\boldsymbol{W}\,,\,\boldsymbol{J}^2\right] = \sum_{\alpha\,\beta} \omega_{\alpha\,\beta}\,\left[\eta_{\alpha\,\beta}\;\boldsymbol{J}^2\right] \tag{111b}$$

As the generators $\eta_{\alpha\beta}$ depend only the cluster, we can collect terms and simplify it to

$$[\boldsymbol{W},\boldsymbol{J}_i] = \boldsymbol{W_{11}}\,[\eta_{11},\boldsymbol{J}_i] + \boldsymbol{W_{22}}\,[\eta_{22},\boldsymbol{J}_i] + 2\boldsymbol{W_{12}}\,[\boldsymbol{J}_x,\boldsymbol{J}_i] \tag{111c}$$

$$\left[\boldsymbol{W},\boldsymbol{J}^2\right] = \boldsymbol{W_{11}}\,\left[\eta_{11},\boldsymbol{J}^2\right] + \boldsymbol{W_{22}}\,\left[\eta_{22},\boldsymbol{J}^2\right] + 2\boldsymbol{W_{12}}\,\left[\boldsymbol{J}_x,\boldsymbol{J}^2\right] \tag{111d}$$

Plugging in Eqs.(107), it is straightforward to see that

$$\left[\boldsymbol{W}\,,\boldsymbol{J}^2\right] = 0 \tag{111e}$$

$$[\boldsymbol{W}\,,\boldsymbol{J}_x] = \;\; i\;(\boldsymbol{W}_{11} - \boldsymbol{W}_{22})\;\boldsymbol{J}_y \tag{111f}$$

$$[\boldsymbol{W}\,,\boldsymbol{J}_y] = -i\;(\boldsymbol{W}_{11} - \boldsymbol{W}_{22})\;\boldsymbol{J}_x + \boldsymbol{J}_z\,\boldsymbol{W}_{12} \tag{111g}$$

$$[\boldsymbol{W}\,,\boldsymbol{J}_z] = -i\,\boldsymbol{J}_y\,\boldsymbol{W}_{12} \tag{111h}$$

Therefore, a fully $\mathfrak{sl}(2,\mathbb{C})$ symmetric weight distribution requires that I) the total sum of intercluster weights be zero and II) all intra-cluster partial sums be equal. We also see that the common requirement for an optimal partition of vanishing intercluster weight entails the rotational symmetry around axis z. The additional conserved quantity (besides the total number of elements) is the z-component of spin (label: up/down) assigned to each cluster.

In other words, Eq.(111h) puts on formal grounds the connection between the intuitive criterion of vanishingly small inter-cluster weight and the possibility of discerning between two different clusters.

For the case of partitions with three clusters we can proceed in an analogous way. $\mathfrak{su}(3)$ has 8 generators G_i. We are interested in the commutators of them and the weight distribution tensor:

$$\begin{aligned}[\boldsymbol{W}\,,\boldsymbol{G}_i] \;&=\; \boldsymbol{W_{11}}\,[\eta_{11},\boldsymbol{G}_i] + \boldsymbol{W_{22}}\,[\eta_{22},\boldsymbol{G}_i] + \boldsymbol{W_{33}}\,[\eta_{33},\boldsymbol{G}_i] \\ &+ \sum_{l<m=1}^{3} \boldsymbol{W_{lm}}\,[(\boldsymbol{\eta}_{lm} + \boldsymbol{\eta}_{ml})\,,\boldsymbol{G}_i]\end{aligned} \tag{112}$$

VI. STATISTICAL MECHANICS

A. On a microcanonical description

Consider the problem of clustering a set of N homeodomain protein sequences. As discussed in section II, even if we are given just with N elements to cluster, we can always consider such a system as a sample of a much larger set N'. For the combined set, we can then write the affinity matrix as

$$\boldsymbol{w} \;=\; \left(\begin{array}{c|c} \boldsymbol{w}_B & \boldsymbol{w}_{BS} \\ \hline & \boldsymbol{w}_S \end{array}\right)$$

and the total weight can be split as sum of the total weight of our system S, that of the additional larger set B and that of their correlation (interaction) as

$$\omega \;=\; \omega_S \;+\; \omega_B \;+\; \omega_{BS}. \tag{113}$$

When considering only the set (system) S of N sequences, B is akin to the *rest of the universe* (of homeodomains) within which the former has been isolated. That is, there has been a process of identifying S as distinct from the rest (clustering), and a step of *erasure* of (that) information and its substitution by an uncertainty Δ in both ω_B and ω_S. The rationale is as follows: Differences in the curation step while gathering the initial data $T = B + S$, as well as differences in the clustering step, will mean that, if we have two observers independently doing this work, both will end up with a different set of sequences (nodes), and hence affinities (i.e., edges contributing to ω_B and ω_S) as well, for both subsystems B and S. If the initial data allows for a clear distinction of both subsets and the accuracy of both procedures is similar, we expect the difference Δ in total weight ω to be small. In the case when

$$|\omega_{BS}| \;\ll\; |\omega_S|,\, |\omega_B| \tag{114}$$

both subsets B and S are weakly correlated and we will also denote such a case as that of two weakly *interacting* systems. If in addition we have $N_B \gg N$ and

$$|\omega_{BS}| \;\ll\; |\omega_S| \;\ll\; |\omega_B|, \tag{115}$$

we may consider the system B as akin to the concept of *heat bath* in statistical mechanics.

Let us consider a fixed topology of the network $\Xi(\omega_{\alpha\beta})$. In general, the number of affinity configurations (*states*) $\boldsymbol{\omega}$ of the system $T = B + S$ compatible with ω is a function of ω and $N' = N_B + N$, $\Omega_T(\omega, N')$. However, this is no longer true when B is acting as a heat bath for S as, by definition, in such a case the condition (115) must hold. In particular, this condition will rule out all those distribution of affinity values for which $|\omega_{BS}| \lesssim |\omega_S|$ {MAKES SENSE?}.

The following description closely follows the arguments in [60]. What are the number of *states* Ω compatible with a given value of ω? The set of states compatible with ω forms a continuous $d_{\boldsymbol{\omega}} \equiv N(N-1)/2 - 1$ dimensional hyperplane according to (88) and (89)

$$0 \;=\; \sum_{\alpha<\beta} d\omega_{\alpha\beta}. \tag{116}$$

The number of states with a total weight at most ω, for large ω, is proportional to $\omega^{d_{\boldsymbol{\omega}}}$, so that the number of states per unit weight range is $\eta(\omega) \propto (d/d\omega)\omega^{d_{\boldsymbol{\omega}}} \propto \omega^{d_{\boldsymbol{\omega}}-1}$. We would like to identify

$$\beta \equiv \frac{d\log\eta(\omega)}{d\omega} = \frac{d_{\boldsymbol{\omega}}-1}{\omega} \tag{117}$$

with the parameter that controls the probability of having the system with total weight ω

$$\mathcal{P}(\omega) \propto e^{-\beta\omega}, \tag{118}$$

the dependence on Γ_N being included in the proportionality constant. Implicitly, it is assumed that there is no absolute scale for the weights $\boldsymbol{\omega}_{ij}$, but only weight differences matter.

Clearly, a description like (118) does not allow us to infer the most likely clustering out of the information contained in $\boldsymbol{\omega}$, as it is $\Omega = \Omega(\omega, N)$. Rather, what we are looking for is a description $\mathcal{P}(\boldsymbol{\omega}, \boldsymbol{a})$ that contains both the information on the graph as well as on the partition. Such a description could convey the intuitive idea that not all partitions fit equally well a given affinity graph $\boldsymbol{\omega}$.

We define the *ideal partition* $\boldsymbol{a}^E$ (or *equilibrium* partition) as one for which the inter-cluster weights are negligible in comparison to the intra-cluster weights, i.e.,

$$W = \frac{1}{2}\sum_i \boldsymbol{W}_{ii} + \sum_{i<j}\boldsymbol{W}_{ij} \simeq \frac{1}{2}\sum_i \boldsymbol{W}_{ii} + \cdots \tag{119}$$

Here *ideal* is meant as an idealized simple solution much like in *ideal gas* in physics, and *equilibrium* is meant as the *true* solution.

For such a description, the details of the partition $\boldsymbol{a}$ will become essential. From the definition, the system is composed of $\mathcal{K} \equiv \|P\|$ independent and *non-interacting* clusters. Thus

$$W^E \simeq \sum_i W_i = \frac{1}{2}Tr(\boldsymbol{a}^{E\dagger}\boldsymbol{\omega}\boldsymbol{a}^E). \tag{120}$$

Swapping elements between clusters, in general, will change the total weight W^E by way of adding a term of inter-cluster interaction

$$W^E \to \sum_i W'_i + \sum_{kl}\boldsymbol{W}'_{kl}. \tag{121}$$

Now, the set of *microstates* compatible with a given W^E is no longer the whole space Π_N. Most importantly, now it is $\Omega = \Omega(\boldsymbol{\omega}, \boldsymbol{a}^E)$, where the dependency on N is already included in $\boldsymbol{\omega}$ and $\boldsymbol{a}^E$.

Appendix A: Non-information theoretic distances

Other distances can also be defined, e.g., the van Dongen distance used in MCL, which we will not discuss in this work. Another interesting distance is the Tarantola distance [28], for which we need to introduce the corresponding Jeffrey function for a partition $J(P)$. This should be a unique, non-negative defined function for each partition. The Tarantola distance between two partitions P and Q is then defined as

$$d_s(P,Q) \equiv \left|\log\frac{J(P)}{J(Q)}\right| \tag{A1}$$

Alternatively, we may introduce a distance based on the rotation matrix $\boldsymbol{U}(P)$ associated to a partition P (Eq. 62).

$$d_m(P,Q) \equiv \left\|\log\frac{\boldsymbol{U}(P)}{\boldsymbol{U}(Q)}\right\|, \tag{A2}$$

where the norm of a matrix is defined as $\|\boldsymbol{U}\|^2 = \frac{1}{2}Tr\,\boldsymbol{U}^\dagger\boldsymbol{U}$. This is a sound definition of distance as $\boldsymbol{U}$ is a symmetric, positive-definite matrix.

It is important to remark that such a distance cannot be defined using the adjacency matrix $\boldsymbol{A}(P)$ of P, as in general it is not invertible -that is true, for instance, whenever there are no singletons. The same applies for any undirected graph $\boldsymbol{\omega}$ as, in general, it is a singular matrix. However, we can consider the traceless part of $\boldsymbol{A}(p)$ given by $\boldsymbol{A}_0 \equiv \boldsymbol{A} - \boldsymbol{1}_{N\times N}$. Analogously, we can introduce a *topological distance* by applying definition (A2) to the adjacency matrix $\mathcal{A}_\omega$.

Provided with such distance $d_m(P,Q)$, the space of (fuzzy) partitions acquires a structure that is not necessarily that of a flat (Euclidean) space[28]. A geodesic, relative to the metric $d_m(P,Q)$, can then be defined joining any pair of partitions P and Q. An interesting question then arises:

Q 13. *is there* always *a geodesic between P and Q that includes partition $P \wedge Q$? If that would be the case, we could write without loss of generality $d_m(P,Q) = d_m(P, P\wedge Q) + d_m(P\wedge Q, Q)$.*

as well as

Q 14. *Can we define a non-information theoretic distance for which its associate geodesics always include $P \wedge Q$?*

Being a rotation, we can always write $\boldsymbol{U}$ in exponential form as $U(P) = e^{i\,\vec{\theta_P}\cdot\vec{\boldsymbol{J}}}$, where $\theta^i_P \in \mathbb{R}$ and $\boldsymbol{J}^i$ are skew-symmetric matrices that form the associated Lie algebra $\mathfrak{su}(N)$. Taking the logarithm we can write it as $\log\boldsymbol{U} = i\,\vec{\theta_P}\cdot\vec{\boldsymbol{J}}$, where the logarithm will be given by

$$\log\boldsymbol{U} = \boldsymbol{C}\,\log\hat{\boldsymbol{U}}\,\boldsymbol{C}^\dagger \tag{A3a}$$

where $\hat{\boldsymbol{U}}$ is the diagonal matrix of the eigenvalues of $\boldsymbol{U}$ and $\boldsymbol{C} \in SU(N)$ is a unitary matrix. Thus, in this representation we will deal with complex matrices $\hat{\boldsymbol{U}}$ and $\boldsymbol{C}$ despite $\boldsymbol{U}$ being real valued.

Let's consider the examples of section IV C, $P_2 = 1|23$ and $P_2'' = 12|3$. For P_2, $\boldsymbol{U}_2$ has eigenvalues $\{1, \lambda_\pm = \frac{(1\pm i)}{\sqrt{2}} = e^{\pm i\,\theta_2}\}$, with $\theta_2 = \pi/4$, and we have

$$\hat{\boldsymbol{U}}_2 \;=\; diag(1, e^{i\,\frac{\pi}{4}}, e^{-i\,\frac{\pi}{4}}) \tag{A3b}$$

$$\boldsymbol{C}_2 \;=\; \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\lambda_+}{\sqrt{2}} & \frac{-\lambda_+}{\sqrt{2}} \\ 0 & \frac{\lambda_-}{\sqrt{2}} & \frac{\lambda_-}{\sqrt{2}} \end{bmatrix}, \tag{A3c}$$

which yields

$$\log \boldsymbol{U}_2 \;=\; \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\theta_2 \\ 0 & \theta_2 & 0 \end{bmatrix} \;=\; -i\,\theta_2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{bmatrix} \;=\; -i\,\theta_2\,\boldsymbol{\lambda_7} \tag{A3d}$$

where $\boldsymbol{\lambda_7}$ is one of the corresponding Gell-Mann matrices that form the Lie algebra of $\mathfrak{su}(3)$. Therefore it is

$$\|\log \boldsymbol{U}_2\| \;=\; \theta_2 \;=\; \frac{\pi}{4}. \tag{A3e}$$

Analogously for P_2'', $\boldsymbol{U}_2''$ has eigenvalues $\{1, \lambda_\pm = \frac{-(2-\sqrt{2})\pm i\,\sqrt{10+4\sqrt{2}}}{4} = e^{\pm\, i(\pi-\theta_2'')}\}$, where $\theta_2'' = \arctan(\frac{\sqrt{10+4\sqrt{2}}}{2-\sqrt{2}})$ and

$$\boldsymbol{C}_2'' \;=\; \begin{bmatrix} \frac{1}{n_1} & \frac{\lambda_+}{n_2} & \frac{\lambda_-}{n_2} \\ \frac{1+\sqrt{2}}{n_1} & \frac{1+\sqrt{2}\lambda_+}{n_2} & \frac{1+\sqrt{2}\lambda_-}{n_2} \\ \frac{1}{n_1} & \frac{1}{n_2} & \frac{1}{n_2} \end{bmatrix}, \tag{A4a}$$

where $n_1 = \sqrt{5+2\sqrt{2}}$ and $n_2 = \sqrt{3-\frac{1}{\sqrt{2}}}$, such that

$$\log \boldsymbol{U}_2'' \;=\; \eta\,(\pi - \theta_2'') \begin{bmatrix} 0 & a & -b \\ -a & 0 & a \\ b & -a & 0 \end{bmatrix}, \tag{A4b}$$

with $a = 7\sqrt{2} - 8$, $b = 6 - \sqrt{2}$ and $\eta = \frac{\sqrt{10+4\sqrt{2}}}{2(19-6\sqrt{2})}$. Its norm is

$$\begin{aligned} \|\log \boldsymbol{U}_2''\| \;&=\; \pi \;-\; \theta_2'' \;= \\ &=\; \pi \;-\; \arctan\left(\frac{\sqrt{10-4\sqrt{2}}}{2-\sqrt{2}}\right) \;\sim\; 0.5468\,\pi. \end{aligned} \tag{A4c}$$

As it is $P_2 \wedge P_2'' = \bar{0}$ and the rotation associated to partition $\bar{0}$ is the identity, we have

$$d_m(P_2, \bar{0}) \;=\; \|\log \boldsymbol{U}_2\| \;=\; \frac{\pi}{4} \tag{A5}$$

$$d_m(P_2'', \bar{0}) \;=\; \|\log \boldsymbol{U}_2''\| \;\sim\; 0.5468\,\pi \;\geq\; \frac{5\pi}{12}. \tag{A6}$$

We want to compare these distances with that between P_2 and P_2''

$$d_m(P_2, P_2'') = \|\log \frac{\boldsymbol{U}_2}{\boldsymbol{U}_2''}\| = \|\log \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\| \;=\; \frac{2\,\pi}{3}. \tag{A7}$$

Appendix B: Metric-generating function

1. Special-strong subadditivity

We define a scalar field $g(P)$ in Π as a function that maps any partition to a real number. For any given scalar field g and all $P, Q, R \in \Pi$, define $\Delta_g\,(P, Q; R) \equiv g(P \wedge R) + g(R \wedge Q) - g(P \wedge Q) - g(R)$. We introduce the following

Definition 1. *We will say that the scalar field g on Π satisfies the special-strong subadditivity (sSA)condition if for all $P, Q, R \in \Pi$ it holds that*

$$\Delta_g\,(P, Q; R) \geq 0. \tag{B1}$$

Below We will justify refering to (B1) as a subadditivity condition. In this section, we will discuss some important consequences of property (B1).

Lema B.1. *If a scalar function g on Π satisfies the special-strong subadditivity condition (sSA), then it holds*

$$P \lhd R \;\Rightarrow\; g(P) \;>\; g(R). \tag{B2}$$

Property (B2), which we may call monotonicity, expresses the compatibility of g with the lattice structure of Π. This follows trivially from (B1) by taking $Q = \bar{0}$ and using (11).

Hence, as Π is a compact lattice, it follows the following

Corolary B.1. *If g satisfies sSA, then it has a minimum and a maximum on Π given by $g(\bar{1})$ and $g(\bar{0})$ respectively.*

We are particularly interested in defining and charectizing a distance metric in the space Π of all partitions.

Definition 2. *We will say that $g(P)$ is a metric-generating function on Π (MGF) when for any P, $Q \in \Pi$ the function $d(P, Q)$ defined as*

$$d(P, Q) \;\equiv\; |\Delta\, g_{P,Q}| \;\equiv\; |g(P) \;-\; g(Q)| \tag{B3}$$

is a pseudometric in Π and the function $D(P, Q)$ defined as

$$D(P, Q) \equiv |g(P \wedge Q) - g(P)| + |g(P \wedge Q) - g(Q)|\,, \tag{B4}$$

which, using the monotonicity property (B2) , can be written as

$$D(P, Q) = \Delta\Delta\, g_{P,Q} \;\equiv\; 2\,g\,(P \;\wedge\; Q) \;-\; g(P) \;-\; g(Q), \tag{B5}$$

is a metric in the space of all partitions.

In that way, each metric-generating function $g(p)$ defines a unique distance in Π. An interesting question is to find the most general conditions for a scalar field g to be a MGF on Π. Here we will address a more modest goal of stating a sufficient condition:

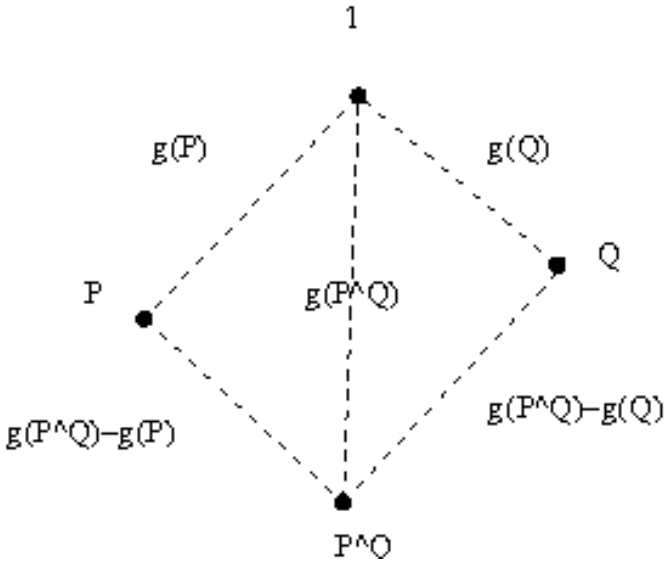


FIG. 7: The distance between two partitions P, Q as sum of edge lengths. Dashed edges are drawn between partitions where one is *smaller* than the other, according to the partial order of the poset. In other words, the lower one is a perfect refinement of the upper one.

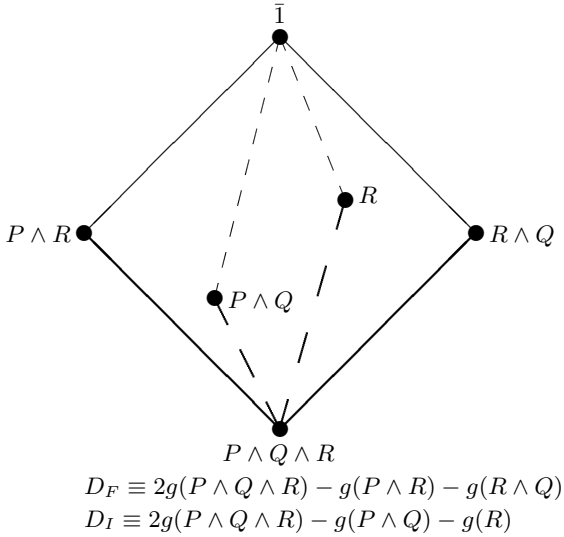


FIG. 8: Hasse Diagram describing condition sSA (B1). Two partitions are joined by a line if (the lower) one is a perfect refinement of the other. The distance between $P \wedge R$ and $R \wedge Q$, D_F, is shown as a continuous bold line, and that between $P \wedge Q$ and R, D_I, as a long-dashed bold line. The question is: is it always $D_I \geq D_F$? The functions g for which that is true are said to satisfy the special-strong subadditivity condition and give rise to the algorithmic complexity (or VI) distance metric in Π.

Lema B.2. *The family of scalar fields g on Π satisfying the special-strong subadditivity condition (B1) is a set of MGFs.*

Proof. Let us prove Lema B.2. We need to show that $D(P,Q)$ is indeed a distance in Π. Clearly, it is a symmetric function of P and Q, i.e., $D(P,Q) = D(Q,P)$ and it is positive defined. Also, from (B2) we have $D(P,Q) = 0 \Leftrightarrow P = Q$. Thus, we only need to demonstrate that $D(P,Q)$ also satisfies the triangular inequality

$$\forall P,Q,R \in \Pi\,;\; D(P,Q) \;\leq\; D(P,R) \;+\; D(R,Q).$$

Inserting B5 into the r.h.s. of this relation, we obtain

$$D(P,R) \;+\; D(R,Q) \;=\; 2\,f(P,Q,R) \;+\; D(P,Q)$$

with $f(P,Q,R)$ such that

$$f(P,Q,R) \equiv -g(P\wedge Q) - g(R) + g(P\wedge R) + g(R\wedge Q) \geq 0$$

if g satisfies the requirement (B1) of a MGF. The distance (B3), however, is only a pseudometric as the MGF g is not required to be an injective function. □

Let us introduce some notation by defining, for all $P,Q,R \in \Pi$,

$$\begin{aligned} g(P|Q) &\equiv g(P\wedge Q) - g(Q) && \text{(B6a)}\\ g(P:Q) &\equiv g(P) + g(Q) - g(P\wedge Q) && \text{(B6b)}\\ g(P:Q|R) &\equiv g(P|R) + g(Q|R) - g(P\wedge Q|R) \\ &= g(P\wedge R) + g(R\wedge Q) - \\ &\quad -g(R) - g(P\wedge Q\wedge R). && \text{(B6c)} \end{aligned}$$

Imposing the non-negativity of function (B6b) is equivalent to the so-called subadditivity condition (SA) $g(P:Q) \geq 0$. What is usually known as strong subadditivity (SSA) condition [61] states the non-negativity of function (B6c), i.e., $g(P:Q|R) \geq 0$. Taking $Q = \bar{0}$, it is straightforward to see that strong-subadditivity implies monotonicity (B2). Using definitions (B6), we can express Δ_g from (B1) as

$$\Delta_g = g(P:Q|R) + g(P\wedge Q\wedge R|P\wedge Q). \qquad \text{(B7)}$$

From the previous discussion we have then that assuming sSA it follows immediately the inequalities

$$g(P|Q) \geq 0 \qquad \text{(B8)}$$

$$g(P:Q) \geq g(\bar{1}), \qquad \text{(B9)}$$

where the first states that the *conditional* g-value of P given Q is positive defined. Monotonicity of g guaranties that the equality is valid iif $P = Q$. The second inequality (B9), which follows from sSA and taking $R = \bar{1}$ in Eq.(B7), corresponds to the SA condition of g if we assume $g(\bar{1}) = 0$.

Hence, condition (B1) is akin to strong subadditivity, in the sense that subadditivity follows immediately from (B1) by taking $R = \bar{1}$ (assuming $g(\bar{1}) \geq 0$), but it is slightly weaker than SSA, as we will see below. This justifies refering to (B1) as a kind of subadditivity condition.

It is important to remark that, in general, however, both conditions SSA and sSA are independent: Clearly, from (B7), it follows that SSA $\Rightarrow$ sSA, but the reverse is not necessarily true. Thus we can state the following

Corolary B.2. *If a scalar function g in Π satisfies $g(\bar{1}) \geq 0$ then we have the following order of implications between the different subadditivity conditions: $SSA \Rightarrow sSA \Rightarrow SA$.*

Monotonicity guaranties the first implication, while the non-negativity of the minimum value of g, that of the second. This result justifies the previous claim of sSA being slightly weaker than SSA, as monotonicity is arguably a sound condition that should be expected from a function defined on Π.

A question that seems worth exploring is whether the special-strong subadditivity sSA applies, and is also relevant, in the context of Quantum Computation:

Q 15. *Does the entanglement entropy satisfies a sSA-equivalent condition? If so, in which cases would we have enough imposing sSA but not SSA?*

2. Shannon entropy as MGF

Particular interesting cases are the following family of information theoretic entropy functions:

$$\begin{aligned}
\text{Rènyi} : g(P) &= H_q(P) \equiv \frac{1}{1-q} \log \sum_k p_k^q \\
\text{Shannon} : g(P) &= H_1(P) = -\sum_k p_k \log p_k \\
\text{Boltzmann} : g(P) &= H_0(P) = \log \|P\| \\
\text{Tsallis} : g(P) &= T_q(P) \equiv \frac{1}{q-1} \left\{ 1 - \sum_k p_k^q \right\} \\
\text{Edit-distance} : g(P) &= T_0(P) = \|P\| - 1
\end{aligned}$$

where $\{p_k\}$ is a probability distribution defined over the clusters $\{P_k\}$ of P. Let us postpone for the moment to clarify the rationale for introducing such a probabilistic description. All these MGFs satisfy that $g(\bar{1}) = 0$. Also, when $\nu \equiv \sum_k p_k^q - 1 \sim 0$ it is $H_q = T_q + O(\nu^2)$.

The Rènyi entropy $H_q(P)$ in $\Pi(S)$ gives rise to a family of distances parametrized by the *extensivity* coefficient q. In the limit $q \to 1$, both the Rènyi and Tsallis entropy coincide with the Shannon entropy associated with a partition P. For the Shannon entropy, we will see that

$$H_1(P \wedge Q) = H_1(P, Q), \tag{B10}$$

i.e., the Shannon entropy of the intersection of P and Q coincides with the (Shannon) joint entropy of P and Q. Therefore, a conditional entropy $H_1(Q|P)$ can be defined and we have the following important identity

$$H_1(Q|P) \equiv H_1(P \wedge Q) - H_1(P) \tag{B11}$$

If we define $p_k = \frac{n_k}{N}$, with $n_k = \|P_k\|$, the Shannon distance associated with $g(P) = H_1(P) = -\sum_k \frac{n_k}{N} \log(\frac{n_k}{N})$ coincides with the so-called VI distance introduced in [32] for partitions:

$$d(P, Q)_{VI} \equiv H_1(Q|P) + H_1(P|Q) \tag{B12}$$

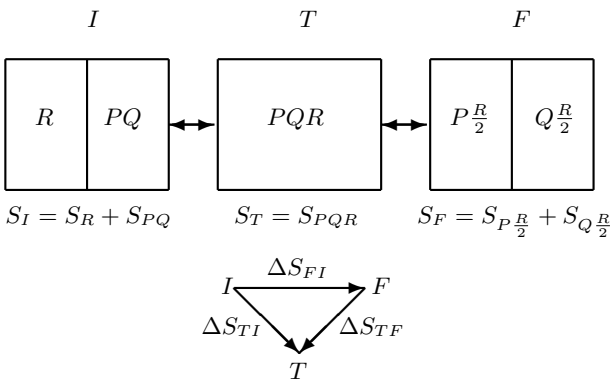


FIG. 9: Entropy of mixing 3 physical components P, Q and R. The box in I contains a pre-mixture of components P and Q which is isolated from component R through an adiabatic wall; F also consists of two isolated subsystems containing the pre-mixtures PZ and ZQ. The amount of molecules of each component, N_P, N_Q and N_R, is the same in both boxes. Removing the internal wall in I or F leads to the mixing of all three components, given by T. The transformation from $I \to F$ has an entropy cost of $\Delta S_{FI} = S_F - S_I$. This transformation can be thought of as a two step process, $I \to T \to F$, through the intermediate state T. Thus, it is $\Delta S_{FI} = \Delta S_{TI} - \Delta S_{TF}$. Is it always $\Delta S_{FI} \geq 0$? We have $\Delta S = k_B \log \Omega$, where $\Omega_{TI} = \frac{N_{PQR}!}{N_P! N_Q! N_R!}$ and $\Omega_{TF} = \frac{N_{PQ}!}{N_P! N_Q!}$, with $N_{PQR} = N_P + N_Q + N_R$. Thus $\Delta S_{FI} = \log \frac{N_{PQR}!}{N_R!} - \log N_{PQ}! = \log \frac{N_{PQR}!}{N_R! N_{PQ}!} \geq 0$.

For the Tsallis entropy, the limit $q \to 0$ yields $T_q \to T_0 \equiv \|P\| - 1$ which defines the Edit distance between two partitions [62].

Appendix C: Formal matrix representation of a partition

Let us first recall some basic definitions of linear algebra. Given a vectorial space $\mathbb{S}$, its *dual* space, denoted as $\mathbb{S}^*$, is a linear space where the vectors (forms) $\mathbf{f}$ are linear functions from $\mathbb{S}$ in to a field $F(= \mathbb{R}, \mathbb{C})$

$$\begin{aligned}
\mathbf{f} &: \mathbb{S} \to F \\
&\mathbf{v} \mapsto \mathbf{f}(\mathbf{v}) \equiv \langle \mathbf{f}, \mathbf{v} \rangle .
\end{aligned} \tag{C1}$$

Let $\{\mathbf{b}_\beta\}$ and $\{\mathbf{b}^\alpha\}$ be arbitrary orthonormal basis of $\mathbb{S}$ and $\mathbb{S}^*$, respectively. We will say that these are dual basis if $\langle \mathbf{b}^\alpha , \mathbf{b}_\beta \rangle = \delta^\alpha{}_\beta$. For a general vector $\mathbf{v}$ of $\mathbb{S}$ and $\mathbf{f}$ of $\mathbb{S}^*$ we can define their components in their respective basis as $\mathbf{v} = v^i \mathbf{b}_i$ and $\mathbf{f} = f_i \mathbf{b}^i$, respectively, where summation over repeated indices is assumed. The linearity property

of forms implies

$$\begin{aligned} \langle \mathbf{f}, \mathbf{v} \rangle &= f_i \, v^i \in F \\ f_i &= \langle \mathbf{f} \, , \, \mathbf{b}_i \rangle \\ v^i &= \left\langle \mathbf{b}^i \, , \, \mathbf{v} \right\rangle . \end{aligned}$$

We will say that $\mathbb{S}$ has a (pseudo) metric if there is a bijective, linear mapping $\mathbf{G}_{\mathbb{S}}$ from $\mathbb{S}$ to $\mathbb{S}^*$ that is also *symmetric*, i.e.,

$$\begin{aligned} \mathbf{G}_{\mathbb{S}} \; &: \; \mathbb{S} \longrightarrow \mathbb{S}^* \\ &\mathbf{v} \mapsto \mathbf{G}_{\mathbb{S}}(\mathbf{v}) \end{aligned} \tag{C2a}$$

and for any two vectors $\mathbf{v}$ and $\mathbf{w}$ of $\mathbb{S}$ it is

$$\langle \mathbf{G}_{\mathbb{S}}(\mathbf{v}) \, , \, \mathbf{w} \rangle \; = \; \langle \mathbf{G}_{\mathbb{S}}(\mathbf{w}) \, , \, \mathbf{v} \rangle \; . \tag{C2b}$$

If $\mathbb{S}$ is endowed with a metric, we can define the *scalar product* of any two vectors $\mathbf{v}$ and $\mathbf{w}$ of $\mathbb{S}$, denoted as, $(\mathbf{v} \, , \, \mathbf{w})$ as

$$(\mathbf{v} \, , \, \mathbf{w}) \; \equiv \; \langle \mathbf{G}_{\mathbb{S}}(\mathbf{v}) \, , \, \mathbf{w} \rangle \; . \tag{C3a}$$

The analogous definition for any two forms $\mathbf{f}$ and $\mathbf{g}$ of $\mathbb{S}^*$ is

$$(\mathbf{f} \, , \, \mathbf{g}) \; \equiv \; \left\langle \mathbf{f} \, , \, {\mathbf{G}_{\mathbb{S}}}^{-1}(\mathbf{g}) \right\rangle \; . \tag{C3b}$$

The *norm* of a vector $\mathbf{v}$ is defined as

$$\| \mathbf{v} \| \equiv \sqrt{(\mathbf{v} \, , \, \mathbf{v})} \, . \tag{C4}$$

We will denote the *entries* of matrix $\mathbf{G}_{\mathbb{S}}$ as $\mathbf{G}_{ij}$, and those of its inverse ${\mathbf{G}_{\mathbb{S}}}^{-1}$, $\mathbf{G}^{ij}$. Hereafter, we will consider an *euclidean* metric, i.e, $\mathbf{G}_{\mathbb{S}} = \mathbf{1}_{N \times N}$, the identity matrix. In terms of its entries it is $G_{ij} = \delta_{ij}$.

Linear representation of a partition: To a set S of N elements $\{s_\alpha\}$ we will assign a *orthonormal* basis $\{\mathbf{e}_\alpha\}$ of vectors in a N-dimensional vectorial space $\mathbb{S}$.

We define a *clustering* on S as a set $\{\mathbf{c}^k\}$ of N, linearly independent forms, *not necessarily* orthogonal, such that for each cluster $\mathbf{c}^k$ it is

$$\begin{aligned} \mathbf{c}^k \; &: \; \mathbb{S} \longrightarrow F \\ &\mathbf{e}_\alpha \mapsto \left\langle \mathbf{c}^k \, , \, \mathbf{e}_\alpha \right\rangle \; = \; {\mu^k}_\alpha \; = \; {\delta^k}_{P(\alpha)}, \end{aligned} \tag{C5a}$$

where the last equality corresponds to the case of a crisp partition and $\{\mathbf{e}_\alpha\}$ is the vectorial basis representing the elements of S. In addition we will require the following *normalization*

$$\sum_k^{\mathcal{K}} \| \mathbf{c}^k \|^2 \; = \; \sum_k^{\mathcal{K}} \mu_\alpha^k \, G^{\alpha\beta} \, \mu_\beta^k \; = \; \sum_k^{\mathcal{K}} \mu_\alpha^k \, \mu^{k\alpha} \; = \; N \, . \tag{C5b}$$

It follows that $0 \leq w_k \equiv \| \mathbf{c}^k \|^2 / N \leq 1$. Probability of cluster k??. For crisp partitions the norm of a cluster is simply the number of elements within that cluster, i.e., $\| \mathbf{c}^k \|^2 = n_k$.

We may endow a clustering with an associated map and its corresponding matrix representation as

$$\begin{aligned} \mathbf{K}^\dagger \; &: \; \mathbb{S}^* \longrightarrow \mathbb{S} \\ &\mathbf{c}^k \mapsto \lambda^k \; = \; \mathbf{K}^\dagger \, \mathbf{c}^k . \end{aligned} \tag{C6}$$

This way we would have to define $\mathbf{K}$ as the adjoint of $\mathbf{K}^\dagger$, namely

$$\begin{aligned} \mathbf{K} \; &: \; \mathbb{S} \longrightarrow \mathbb{S}^* \\ \mathbf{K} \; &= \; \mathbf{G}_{\mathbb{S}} \left(\mathbf{K}^\dagger \right)^\intercal \mathbf{G}_{\mathbb{S}} \end{aligned} \tag{C7}$$

Note that, by definition of adjoint, it is $\mathbf{K}^\dagger \equiv \mathbf{G}_{\mathbb{S}}^{-1} \, \mathbf{K}^\intercal \, \mathbf{G}_{\mathbb{S}^*}$ and with $\mathbf{G}_{\mathbb{S}^*} = \mathbf{G}_{\mathbb{S}}^{-1}$. In general, it is $\mathcal{K} \equiv rank(\mathbf{K}^\dagger) = rank(\mathbf{K}) \leq dim(\mathbb{S}^*) = N$.

The corresponding *equivalence relation* $\mathbf{A}$ is defined as the linear map

$$\mathbf{A} \; = \; \mathbf{K}^\dagger \mathbf{K} \; : \; \mathbb{S} \longrightarrow \mathbb{S}. \tag{C8a}$$

The entries of $\mathbf{A}$ are

$${\mathbf{A}^\alpha}_\beta \; = \; \left\langle \mathbf{e}^\alpha \, , \, \mathbf{K}^\dagger \mathbf{K} \mathbf{e}_\beta \right\rangle . \tag{C8b}$$

Defining the operators

$$\mathbf{Q} \; \equiv \; \mathbf{G}_{\mathbb{S}} \, \mathbf{K}^\dagger \; : \; \mathbb{S}^* \longrightarrow \mathbb{S}^* \tag{C9}$$

$$\mathbf{P} \; \equiv \; \mathbf{K}^\dagger \, \mathbf{G}_{\mathbb{S}} \; : \; \mathbb{S} \longrightarrow \mathbb{S} \tag{C10}$$

it is $\mathbf{P}^\dagger = \left(\mathbf{K}^\dagger \right)^\intercal \mathbf{G}_{\mathbb{S}}$ and $\mathbf{Q}^\dagger = \mathbf{G}_{\mathbb{S}} \left(\mathbf{K}^\dagger \right)^\intercal$, and we may rewrite $\mathbf{A}$ in terms of endomorphisms $\mathbf{P}$ and $\mathbf{P}^\dagger$ as $\mathbf{A} = \mathbf{P} \, \mathbf{P}^\dagger$ and define its complementary operator as

$$\mathbf{N} \equiv \mathbf{Q}^\dagger \mathbf{Q} = \mathbf{K} \mathbf{K}^\dagger . \tag{C11}$$

From their definition, it follows that both operators have the same non-negative eigenvalues $\{\sigma_\alpha^2\}$. We will impose the *normalization* condition

$$Tr \, (\mathbf{A}) \; = \; Tr \, (\mathbf{N}) \; = \; \sum_{i=1}^{N} \sigma_i^2 \; = \; N. \tag{C12}$$

Denoting the singular value decomposition of $\mathbf{K}^\dagger$ as $\mathbf{K}^\dagger = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\dagger$ ($\mathbf{K} = \mathbf{V} \mathbf{\Sigma}^\dagger \mathbf{U}^\dagger$) it is

$$\mathbf{A} \; = \; \mathbf{U} \, \mathbf{\Sigma} \mathbf{\Sigma}^\dagger \, \mathbf{U}^\dagger \tag{C13a}$$

$$\mathbf{N} \; = \; \mathbf{V} \, \mathbf{\Sigma}^\dagger \mathbf{\Sigma} \, \mathbf{V}^\dagger \tag{C13b}$$

$$\mathbf{A} \; = \; \mathbf{R} \, \mathbf{N} \, \mathbf{R}^\dagger \; ; \; \mathbf{R} \equiv \mathbf{U} \mathbf{V}^\dagger , \tag{C13c}$$

such that $\mathbf{R} \mathbf{R}^\dagger = \mathbf{1}$ and $\det(\mathbf{R}) = +1$.

For a crisp partition (61,62), it is $\mathbf{V} = \mathbf{1}$ and $\mathbf{\Sigma} = \hat{\mathbf{a}}$, and we have $\mathbf{K}^\dagger = \mathbf{U} \hat{\mathbf{a}}$. For instance, for the clustering $\{\mathbf{c}^k\} = \{(1,0,0),(0,1,1)\}$ of $N = 3$ elements it is (see

examples of partition rotations below)

$$\mathbf{K}^\dagger = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \mathbf{a} \tag{C14a}$$

$$\mathbf{K} = \hat{\mathbf{a}}\mathbf{U}^\dagger = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{C14b}$$

$$\lambda^1 \equiv \mathbf{K}^\dagger \mathbf{c}^1 = \mathbf{c}^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \tag{C14c}$$

$$\lambda^2 \equiv \mathbf{K}^\dagger \mathbf{c}^2 = \mathbf{c}^2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \tag{C14d}$$

(C14e)

and we see that for a crisp partition the clustering $\{\mathbf{c}^k\}$ constitutes a set of left singular vectors of $\mathbf{K}^\dagger$ with singular value 1, and we recover our initial matrix description of a parition, namely it is

$$\mathbf{K}^\dagger = \left[\cdots \overrightarrow{a}_i \cdots \right] \tag{C15a}$$

$$\mathbf{K} = \left[\cdots \overrightarrow{a}_\alpha \cdots \right] \tag{C15b}$$

with $\overrightarrow{a}_i = 0$ for $i > \mathcal{K}$. Furthermore, each right singular vector $\overrightarrow{d}_i$ $(i = 1, \cdots, \mathcal{K})$ of $\mathbf{K}$ provides a representatitve element of each cluster $\overrightarrow{d}_i = \mathbf{e}_\alpha \, \delta_{iP(\alpha)}$ for a specific element s_α. These last statements require a precision. We claim here that

Conjecture 4. *there is always a reordering of the rows of* $\mathbf{K}$ *(columns of* $\mathbf{K}^\dagger$*) such that* $\{\mathbf{c}^i\}$ *and* $\{\overrightarrow{d}_i\}$ *are manisfestly left and right singular vectors of* $\mathbf{K}^\dagger$*, respectively.*

THOUGHTS: The way we have defined the membership functions and their normalization, however, might not allow for an easy and consistent generalization. An alternative path could lie on realizing that the actual *meaningful observable* might only be the equivalence relation $\mathbf{A}$. Thus, the appropriate *correspondence principle* should be that, for a crisp partition, by definition, it is

$$\langle \mathbf{e}^\alpha \, , \, \mathbf{P}\mathbf{P}^\dagger \, \mathbf{e}_\beta \rangle = \delta^{P(\alpha)}{}_{P(\beta)} \tag{C16a}$$

$$\langle \mathbf{Q}^\dagger \mathbf{Q} \, \mathbf{e}^i \, , \, \mathbf{e}_j \rangle = n_i \, \delta^i{}_j, \tag{C16b}$$

or

$$\left(\mathbf{e}_\alpha \, , \, \mathbf{P}\mathbf{P}^\dagger \, \mathbf{e}_\beta \right) = g_{\alpha\beta} \, \delta_{P(\alpha)P(\beta)} \tag{C16c}$$

$$\left(\mathbf{Q}^\dagger \mathbf{Q} \, \mathbf{e}^i \, , \, \mathbf{e}^j \right) = n_i \, g^{ij}, \tag{C16d}$$

where the cluster size n_i should be defined in terms of $\mathbf{c}^k$ and $\mathbf{K}^\dagger$ in an appropriate and consistent way.

IV. ACKNOWLEDGMENTS

The author would like to thank E. Vives and J.M. Sancho for helpful discussions.

V. NOTES

Notes

[1] **S. Strogatz and D. Watts, Nature 393, 440 (1998).**

[2] **A. Barabási and R. Albert, Science 286, 509 (1999).**

[3] **M. Girvan and M. Newman, PNAS 99, 7821 (2002).**

[4] **S. Fortunato, Phys. Rep. 486, 75 (2010).**

[5] **T. Evans, J. Stat. Mech. P12037 (2010).**

[6] **M. Porter, J. Onnela, and P. Mucha, Not. Amer. Math. Soc. 56, 1082 (2009).**

[7] **S. Schaeffer, Comp. Sci. Rev. 1, 27 (2007).**

[8] **A. Griffiths, H. C. Luckhurst, and P. Willett, J. American Society for Information Science 37, 3–11 (1986).**

[9] **L. Azzopardi, G. Kazai, S. Robertson, S. Rüger, M. Shokouhi, D. Song, and E. Tilmaz, eds., *Second International Conference on the Theory of Information Retrieval* (Springer-Verlag, 2009).**

[10] **R. Justo and I. Torres, in *Progress in Pattern Recognition, Image Analysis and Applications*, edited by M. Lazo and A. Sanfeliu (2005), vol. 3773, pp. 556–565.**

[11] **A. Clark, Ph.D. thesis, University of Sussex (2001), URL http://arxiv.org/abs/cs/0212024.**

[12] **A. Ben-Naim, Journal of Chemical Physics 135, 085104 (2011).**

[13] **S. van Dongen, Ph.D. thesis, University of Utrecht (2000), URL http://micans.org/mcl/lit/svdthesis.pdf.gz.**

[14] **B. Frey and D. Dueck, Science 315, 972 (2007).**

[15] **G. Lance and W. Williams, Computer Journal 9, 373 (1967).**

[16] **G. Lance and W. Williams, Computer Journal 10, 271 (1967).**

[17] **J. Reichardt and S. Bornholdt, Physical Review E 74, 016110 (2006).**

[18] **C. Massen and J. Doyen, *Thermodynamics of community structure* (2006), URL http://arxiv.**

org/abs/cond-mat/0610077v1.
[19] M. Newman and M. Girvan, Physical Review E 69, 026113 (2004).
[20] L. Danon et al., J. Stat. Mech. 09, P09008 (2005).
[21] S. Fortunato and M. Barthelemy, PNAS 104, 36 (2007).
[22] B. Good, Y. de Montjoye, and A. Clauset, Phys. Rev. E 81, 046106 (2010).
[23] K. H. Knuth, in *Proceedings of the 23rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by P. Goggans and C. Chan (American Institute of Physics, 2004), vol. 707, pp. 132–144.
[24] P. Goyal, K. H. Knuth, and J. Skilling, Phys. Rev. A 81, 022109 (2010).
[25] P. Goyal and K. H. Knuth, Symmetry 3, 171 (2011).
[26] D. Ellerman, Synthese 168, 119 (2009).
[27] D. Ellerman, Review of Symbolic Logic 3, 287 (2010).
[28] A. Tarantola, *Elements for Physics* (Springer-Verlag, 2006).
[29] E. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 1996).
[30] W. H. Zurek, Nature 341, 119 (1989).
[31] W. H. Zurek, Physical Review A 40, 4731 (1989).
[32] M. Meilă, in *Proc. 16th Ann. Conf. Computational Learning Theory* (2003), pp. 173–187.
[33] M. Cross and P. Hohenberg, Rev. Mod. Phys. 65, 857 (1993).
[34] J. García-Ojalvo and J. M. Sancho, *Noise in Spatially Extended Systems* (Springer-Verlag, 1999).
[35] R. Defay and I. Prigogine, *Tension Superficielle et Adsorption* (Desoer, 1951).
[36] J. Sakurai, *Modern quatum mechanics* (Addison-Wesley Publishing Company, 1994).
[37] G. E. Andrews, *Number Theory* (Dover Publications, 1994).
[38] M. Maggiore, *A Modern Introduction to Quantum Field Theory* (Oxford University Press, 2005).
[39] P. Chaikin and T. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, 1995).
[40] P.-G. de Gennes and J. Prost, *The Physics of Liquid Crystals* (Oxford University Press, 1995).
[41] S. Weinberg, *The Quantum Theory of Fields* (Cambridge University Press, 1995).
[42] J. Fuchs and C. Schweigert, *Symmetries, Lie Algebras and Representations* (Cambridge University Press, 1997).
[43] J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley Publishing Co., 1994).
[44] M. I. Petrashen and E. D. Trifonov, *Applications of Group Theory in Quantum Mechanics* (Dover Publications Inc., 1969).
[45] A. de Shalit and I. Talmi, *Nuclear Shell Theory* (Dover Publications Inc., 2004).
[46] H. J. Lipkin, *Lie Groups for Pedestrians* (North-Holland Publishing Co., 1965).
[47] E. Jaynes, Rev. Mod. Phys. 106, 620 (1955).
[48] E. Jaynes, Phys. Rev. 108, 171 (1957).
[49] C. H. Bennet, P. Gács, M. Li, P. M. B. Vitányi, and W. H. Zurek, IEEE Transactions on Information Theory 44, 1407 (1998).
[50] G. Ruppeiner, Phys. Rev. A 20, 1608 (1979).
[51] G. Ruppeiner, Rev. Mod. Phys. 63, 605 (1995).
[52] S. Burris and H. Sankappanavar, *A course in Universal Algebra* (Springer-Verlag, 1981).
[53] B. Davey and H. Priestley, *Introduction to Lattices and Order* (Cambridge University Press, 2002).
[54] A. Strehl and J. Ghosh, Journal of Machine Learning Research 3, 583 (2002).
[55] D. Hestenes and G. Sobcyk, *Clifford Algebra to Geometric Calculus: A unified Language for Mathematics and Physics* (Reidel Dordrecht, 1984).
[56] A. Okounkov, *The uses of random partitions* (2003), URL http://arxiv.org/abs/math-ph/0309015v1.
[57] R. Dijkgraaf, D. Orlando, and S. Reffert, *Quantum crystals and spin chains* (2008), URL http://arxiv.org/abs/0803.1927v2.
[58] M. Blatt, S. Wiseman, and E. Domany, Phys. Rev. Let. 76, 3251 (1996).
[59] S. Wiseman, M. Blatt, and E. Domany, Phys. Rev. E 57, 3767 (1998).
[60] R. P. Feynman, *Statistical Mechanics: A Set of lectures* (Leninogorski Pedagoguicheski Institute, 1998), 1982.
[61] D. W. Robinson and D. Ruelle, Communications in Mathematical Physics 5, 288 (1967).
[62] D. Brown and K. Sjoelander, PLOS Comput. Biol 2, 0479 (2007).