



**HAL**  
open science

## Independence versus Indetermination: basis of two canonical clustering criteria

Pierre Jean-Claude Robert Bertrand, Michel Broniatowski, Jean-François Marcotorchino

► **To cite this version:**

Pierre Jean-Claude Robert Bertrand, Michel Broniatowski, Jean-François Marcotorchino. Independence versus Indetermination: basis of two canonical clustering criteria. *Advances in Data Analysis and Classification*, In press, 16 (4), 10.1007/s11634-021-00484-1 . hal-02901167v3

**HAL Id: hal-02901167**

**<https://hal.science/hal-02901167v3>**

Submitted on 4 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Independence versus Indetermination: basis of two canonical clustering criteria

Pierre Bertrand · Michel Broniatowski ·  
Jean-François Marcotorchino

Received: date / Accepted: date

**Abstract** This paper aims at comparing two coupling approaches as basic layers for building clustering criteria, suited for modularizing and clustering very large networks.

We briefly use "optimal transport theory" as a starting point, and a way as well, to derive two canonical couplings: "statistical independence" and "logical indetermination". A symmetric list of properties is provided and notably the so called "Monge's properties", applied to contingency matrices, and justifying the  $\otimes$  versus  $\oplus$  notation. A study is proposed, highlighting "logical indetermination", because it is, by far, lesser known.

Eventually we estimate the average difference between both couplings as the key explanation of their usually close results in network clustering.

**Keywords** Correlation Clustering · Mathematical Relational Analysis · Logical Indetermination · Coupling Functions · Optimal Transport · Graph Theoretical Approaches

## 1 Introduction

Network clustering (or clique-partitioning of graphs) is a key topic, concerned with a very large dedicated literature. One of the reasons of this status is the

---

Pierre Bertrand  
Laboratoire de Probabilités, Statistique et Modélisation — CNRS UMR 8001, Sorbonne Université, Paris, France  
E-mail: pierre.bertrand@ens-paris-saclay.fr

Michel Broniatowski  
Laboratoire de Probabilités, Statistique et Modélisation — CNRS UMR 8001, Sorbonne Université, Paris, France  
E-mail: michel.broniatowski@sorbonne-universite.fr

Jean-François Marcotorchino  
Institut de statistique, Sorbonne Université, Paris, France  
E-mail: jfmmarco3@gmail.com

recent use made by the GAFAM companies about very large networks resulting of modern activities dealing with: *big social networks, cellphone communications networks, high speed financial trading, large IT networks, IOT networks, etc.*. This is simultaneously associated with the IT capacity afforded today to store the really huge amounts of data, those activities force us to cope with. The sudden apparition of these big networks gave rise to a renewal of the so-called *graph theoretical domain*, used in that context for different purposes, such as: discovering the latent cliques, clustering the whole network, isolating some key parts of interest within the network, etc. In other words, this massive and raw information contained inside the networks must be analyzed *per se*, and this leads obviously to mandatory techniques, among which networks clustering plays a prominent role, with a lot of practical contextual applications.

In the scientific literature, it appears that many different methods have been dedicated to graph clustering, one can find in [14] or more recently in [13] a quite interesting overview on this matter. Most of them use a local criterion based on the number of paths [20], the number of shortest paths [17] or a proportion of present edges [27] which is then aggregated to define a global criterion to be optimized. Some methods are based on pure decomposition of a graph, as for instance in [3] where the authors construct a  $k$ -clustering based on a spanning tree by removing  $k - 1$  edges, while some others are concerned directly with spectral analysis of graph laplacians, or with mathematical relational analysis (correlation clustering). In addition to this relatively mainstream list, some methods rely on the application of very specific mathematical domains as those typically addressed through Mean Field Games theory in [11] or more promising, such as the approach given in [29] where a method based on the evolution of a discrete Ricci curvature flow was proposed.

This method deserves to be briefly exposed: given two measures  $\mu$  and  $\nu$  in a space endorsed with a distance  $d$ , the Wasserstein's transportation distance  $W(\mu, \nu)$  is the minimum total weight to move  $\mu$  to  $\nu$  according to  $d$  as presented in [30]. A measure  $m_x^{\alpha,p}$  to capture the neighborhood of a node  $x$  is formulated in [29]. With  $d(x, y)$  the shortest path between two nodes  $x$  and  $y$  in the network, Ricci's curvature then basically expresses whether  $x$  is closer from  $y$  using  $d$  or  $W$  between  $m_x^{\alpha,p}$  and  $m_y^{\alpha,p}$ . Eventually, they are interested in the Ricci Flow (see [18]) which solves a differential equation where the derivative is (almost) this Ricci curvature; consequently if the neighborhood associated to  $x$  and  $y$  and compared by  $W$  are closer than  $x$  and  $y$  compared by  $d$ , the derivative is strictly less than one and conversely. Iteratively updating a weight to solve the Ricci Flow differential equation, before cutting links greater than a threshold, they deliver results rather similar or slightly better on the usual experimental networks. Nevertheless, the lack of a motivated choice for the underlying parameters ( $\alpha$ ,  $p$  and the threshold) is a limitation similar to the choice of a criterion in the usual Louvain Algorithm. This drawback precisely motivates the present paper. Actually, any comparison with competitive approaches would imply a self adaptive procedure to select the technical parameters.

At that stage two aspects must be differentiated: on the one hand (i) the existence of generic algorithms to optimize a *clustering criterion* as a global objective function, or, on the other hand (ii), the construction of network clustering criteria themselves.

As for the first point, (i) concerned with generic algorithms, it is well known that several methods were introduced to fit this purpose and notably the famous Louvain algorithm, whose origin is rather recent [6], and which is recognized as a very good tool by the scientific community. It is originally based upon the optimization of a global function called *modularity* initially defined in [28] and that the community has adopted to compare two clusterings on a common basis. In the sequel we shall denote  $\mathcal{M}^\times$  this objective function and will show it measures *deviation from statistical independence*.

To fulfill the (ii) objective, the Louvain algorithm has been naturally generalized in [8] where the authors proposed to choose a candidate criterion among a list of global criteria, different from the usual *modularity*. Actually, the modularity, due to a resolution limit first mentioned in [15], has been modified in several articles ([32], [21] and [9]) always motivated by experimental results; we will not detail further the list of available criteria. In her thesis [10], Patricia Conde-Céspedes, proposed some experiments on usual networks, involving  $\mathcal{M}^\times$  plus some others, showing that results may vary from one criterion to another, while being still consistent and interpretable.

In this paper, we will focus, on two network clustering criteria she applied, the original  $\mathcal{M}^\times$  and a second denoted  $\mathcal{M}^+$  which is locally based on a deviation to another coupling function, already latent in a paper of Fréchet [16] and that we shall call *indetermination* or *logical indetermination* (a notion introduced by J.F. Marcotorchino in his seminal papers [22] and [24]).

The innovation of this paper can be stated as follows:

- We rely on a work by Csiszár on divergences [12] which assesses that the costs in a projection problem is restricted to Least Squares or Entropy. Leveraging on it we show that the two chosen criteria  $\mathcal{M}^\times$  and  $\mathcal{M}^+$  precisely result from the optimization of the two corresponding canonical discrete transportation problems.
- We gather known and new properties of the so-called *indetermination*, an equilibrium already applied in the graph clustering domain but never studied *per se* in a more general context. Furthermore, the expected difference between the two canonical coupling functions is shown to be of order  $\mathcal{O}(\frac{1}{n^2})$  where  $n$  is the number of nodes.
- We validate some of these findings, by reanalyzing more systematically the behavior of those criteria on the very simple model of Gilbert’s graphs. This last item illustrates the small expected difference of the previous item and explains the close experimental results found in [8]. Besides it motivates the search of situations where the two criteria range significantly apart as briefly considered in the paper.

The paper is structured as follows. In section 2 we propose a parallel discovery of two coupling functions ( $\otimes$ ) and ( $\oplus$ ) using discrete *optimal transport*

*theory*. In section 3 is mentioned a list of dual properties related to Monge's matrices and which justify the notation  $\oplus/\otimes$  that we propose. Section 4 deeply studies *indetermination* introducing properties that, to our knowledge, deserve to be put forward with regards to the too poor coverage devoted to them in the literature. Finally, Section 5 gathers a study about the behavior of the criteria based on those coupling functions on the general Gilbert's random network model. We show they are overall quite similar, which illustrates their symmetric construction.

## 2 Parallel discovery of two dual couplings

When we want to couple two marginal laws, the most common and straightforward way to proceed, consists in assuming *independence* as a handy hypothesis. For instance when we use a very classical and usual criterion like the  $\chi^2$  index, we are measuring nothing but a deviation to *independence*, a natural coupling for empirical experiments when one wants to quantify dependence.

Although being the most natural, it is not, by far, the only existing available coupling method; actually, as introduced by Sklar in [33], any copula function will lead to a coupling function acting on two cumulative distribution functions. In this document, we link a coupling function to a given optimal transport problem. Hence, to follow a similar approach for *indetermination* coupling, we train ourselves first by extracting the *independence* coupling from the optimization of a transport problem and, next generalize the same approach to the *indetermination* case, making use of a second and different transport problem.

We already introduced the term "coupling function" several times but let us define it formally, since it will be a key notion throughout the document.

### Definition 1 (Coupling function)

Let  $\mu = \mu_1 \dots \mu_p$  and  $\nu = \nu_1 \dots \nu_q$  be two discrete probabilities called marginal distributions (or simply margins). A probability measure  $\pi = \pi_{u,v}$   $\{1 \leq u \leq p, 1 \leq v \leq q\}$  on the product space is defined through a coupling function  $C$  such that  $\pi = C(\mu, \nu)$ , satisfying the following constraints:

- (first margin)  $C(\mu, \nu)_{u,\cdot} = \sum_{v=1}^q C(\mu, \nu)_{u,v} = \mu_u, \forall 1 \leq u \leq p$
- (second margin)  $C(\mu, \nu)_{\cdot,v} = \sum_{u=1}^p C(\mu, \nu)_{u,v} = \nu_v, \forall 1 \leq v \leq q$
- (positivity)  $C(\mu, \nu)_{u,v} \geq 0, \forall 1 \leq u \leq p, \forall 1 \leq v \leq q$

### Remark 1

All coupling functions (or maps) used in this paper will satisfy:  $\pi_{u,v} = C(\mu, \nu)_{u,v} = C(\mu_u, \nu_v)$ ; this illustrates that for all  $(u, v)$ ,  $\pi_{u,v}$  only depends upon the value on the corresponding margins:  $\mu_u$  and  $\nu_v$ .

## 2.1 A few words about Optimal Transport

Looking at Definition 1, we observe that a coupling function behaves as a copula in the discrete domain: acting on margins it derives a probability distribution on the product space.

We can imagine a lot of coupling functions, especially if we do not limit ourselves to Remark 1. The constraints that  $C$  has to satisfy, lead us to cope with a huge number of solutions but without any general constructive method. This is the reason why we shall choose a systematic approach: minimizing a cost function and observe the link to *optimal transport* definition.

The ad-hoc discrete optimal transport problem we will be dealing with, typically looks like Problem 1, given hereafter (where MKP stands for Monge-Kantorovitch-Problem).

### Problem 1 (Discrete Version of MKP)

$$\begin{aligned} & \min_{\pi} \sum_{u=1}^p \sum_{v=1}^q \mathbf{C}(\pi(u, v)) \\ \text{subject to:} & \\ & \sum_{v=1}^q \pi(u, v) = \mu_u; \forall u \in \{1, \dots, p\} \\ & \sum_{u=1}^p \pi(u, v) = \nu_v; \forall v \in \{1, \dots, q\} \\ & \pi(u, v) \geq 0; \forall (u, v) \in \{1, \dots, p\} \times \{1, \dots, q\} \end{aligned}$$

The choice of a cost function  $\mathbf{C}$  depends upon the applications we want to address. Typically, we expect the global assignment to be as smooth as possible, meaning close to uniform (see both examples in the sequel). A MKP problem is then essentially given by its cost function, while margins  $(\mu, \nu)$  may vary. Let us define now an optimal coupling function  $C$  associated to a given MKP problem with fixed margins given as parameters.

### Definition 2 (MKP Problem Associated with Coupling function)

For a given MKP problem  $P$ , we define a coupling function  $C^P$  by:  $C^P(\mu, \nu) = \pi^*(P)$  provided that  $\pi^*$  exists as a unique solution of  $P$  with margins  $\mu$  and  $\nu$ .

Following Definition 2 we propose the solutions of two discrete optimal transport problems that we apply in section 5: each signals a structured and well-defined criterion, suitable for network clustering.

## 2.2 The Alan Wilson's Entropy Model: role of "independence"

First introduced by Sir Alan Wilson in 1969 for "Spatial Interaction Modeling" the "Flow Entropy Model" of Alan Wilson, can be found in his various

publications: originated in [36] and developed in [37]. A fundamental justification of his approach corresponds to the following contextual situation: in a theoretical system, the elements of which do not maintain affinities, it is advisable to determine the distribution of  $\pi(u, v)$  (normalized frequency flows), supposing  $\pi \geq 0$  which maximizes the entropy of the system under certain constraints. The objective function to be minimized is based upon *Boltzmann's or Shannon's Entropies* so that the problem should be expressed as follows:

**Problem 2 (Unbalanced PSIS)**

$$\min_{\pi} \sum_{u=1}^p \sum_{v=1}^q \pi(u, v) \ln(\pi(u, v))$$

In a situation with no information, the minimization of Problem 2 just amounts to satisfy the constraint that the cell values distribution is effectively a probability (i.e.: the sum of positive  $\pi(u, v)$  is equal to 1). The solution of this very simple "Program of Spatial Interaction System" (PSIS) is nothing but the uniform law:

$$\pi^*(u, v) = \frac{1}{pq} \quad (1)$$

In other words, when we ignore everything about the way the exchanges are built up, it is necessary to use Laplace's principle of "insufficient reason" and to consider that the world trade is uniformly distributed inside the system.

By using margins, let us say information about total exports (origins flows) and total imports (destination flows), the degree of disorder of the system can be drastically reduced. Indeed, totals on rows and columns are no longer free, but must satisfy marginal values  $\mu_u$  and  $\nu_v$ , fixed by the application as expressed in Problem 3; the solution of which is given by theorem 1.

**Problem 3 (Balanced PSIS)**

$$\min_{\pi} \sum_{u=1}^p \sum_{v=1}^q \pi(u, v) \ln(\pi(u, v))$$

subject to constraints of Problem 1

**Theorem 1**

*The solution of Problem 3 is  $\pi^\times(u, v) = \mu_u \nu_v$ .*

*Hence the coupling function associated to Problem 3 is nothing but "independence":*

$$C^{\text{Problem 3}}(\mu, \nu)_{u,v} = C^\times(\mu, \nu)_{u,v} = (\mu \otimes \nu)_{u,v} = \mu_u \nu_v$$

We skip the proof of Theorem 1 as it is similar to the one we will develop for Theorem 2 which is less common.

As a conclusion, from the direct maximization of entropy, we get the solution expressed in terms of probability and remark that the associated coupling function is nothing but "independence" (expressed with a  $\otimes$  throughout the paper).

### 2.3 The minimal trade model: role of "indetermination"

In the "Minimal Trade Model" (see [35], [22] and [24]), the cost function aims at getting a smooth breakdown of the origins-destinations  $\pi(u, v) = \frac{n_{u,v}}{n_{..}}$  which explains the term "Minimal Trade". In that case the criterion is a quadratic function measuring the squared deviation of the cell values from the "no information" situation (the uniform joint distribution law related to Problem 2). Obviously, in case of free margins, the solution remains the uniform law. However, adding usual pre-conditioned constraints on margins, the least squared problem is Problem 4; the solution of which is given by Theorem 2.

#### Problem 4 (Minimal Trade Model)

$$\min_{\pi} \sum_{u,v} \left\{ \pi(u, v) - \frac{1}{pq} \right\}^2$$

subject to constraints of Problem 1

#### Theorem 2

The solution of Problem 4 is  $\pi^+(u, v) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$ .

Hence the coupling function associated to Problem 4 is nothing but "indetermination":

$$C^{Problem\ 4}(\mu, \nu)_{u,v} = C^+(\mu, \nu)_{u,v} = (\mu \oplus \nu)_{u,v} = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$$

A supplementary condition, which is exogenous with regard to the previous model must be added on the margins (which are, by the way, constant values given *a priori*), this condition (see [22]) is a simple inequality which guarantees the positivity of the frequency Matrix  $\pi^*(u, v)$  we are looking for:

$$p \min_u \mu_u + q \min_v \nu_v \geq 1 \quad (2)$$

From now on, we shall consider that Condition (2) applies whatever the breakdown of the  $\mu_u$  and  $\nu_v$  is. Notice that in the "Adjustment to Fixed Margins for a Contingency Table" case, the associated values  $n_{uv}$  must be integers, and therefore making the problem much more complex to solve, relaxation of this integrity constraint leads formally to Problem 4.

#### Remark 2 (Vanishing bias)

By developing the cost function, we obtain an interesting equality to be used later on:

$$\sum_{u,v} \left( \pi(u, v) - \frac{1}{pq} \right)^2 = \left( \sum_{u,v} \pi^2(u, v) \right) - \frac{1}{pq} \quad (3)$$

so that the influence of the constant shift  $\frac{1}{pq}$  in the squared model disappears.



*Proof*

The proof we propose comes directly from [35] and [24]. A generalization of the canonic additive form when we relax hypothesis (2) can be found in the thesis to come [4].

Using equality (3), the Lagrangian function associated with the previous minimization model can be turned into

$$\begin{aligned} \mathbf{L}(\pi, \lambda, \omega, \theta) &= \sum_{u=1}^p \sum_{v=1}^q \pi^2(u, v) + \sum_{u=1}^p \lambda_u \left( \mu_u - \sum_{v=1}^q \pi(u, v) \right) \\ &+ \sum_{v=1}^q \omega_v \left( \nu_v - \sum_{u=1}^p \pi(u, v) \right) - \theta \left( \sum_{u=1}^p \sum_{v=1}^q \pi(u, v) - 1 \right) \end{aligned}$$

Since the function to optimize is convex, the solution is a minimum so that first order conditions apply and we have the following system of equations.

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \pi(u, v)} = 2\pi(u, v) - \lambda_u - \omega_v - \theta = 0 \quad (4)$$

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \lambda_u} = \mu_u - \sum_{v=1}^q \pi(u, v) = 0 \quad (5)$$

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \omega_v} = \nu_v - \sum_{u=1}^p \pi(u, v) = 0 \quad (6)$$

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \theta} = \sum_{u=1}^p \sum_{v=1}^q \pi(u, v) - 1 = 0 \quad (7)$$

Since the Lagrange multipliers are defined up to a constant we may assume  $\theta = 0$ ; summing on  $v$  in equation (4), we obtain

$$2\mu_u = q\lambda_u + \sum_{v=1}^q \omega_v; \quad (8)$$

which rewrites

$$\lambda_u = \frac{2\mu_u}{q} - \frac{\sum_{v=1}^q \omega_v}{q}. \quad (9)$$

Similarly we have:

$$\omega_v = \frac{2\nu_v}{p} - \frac{\sum_{u=1}^p \lambda_u}{p}. \quad (10)$$

Introducing equation (9) and equation (10) in equation (4) we get:

$$\pi_{u,v} = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{\sum_{u=1}^p \lambda_u}{2p} - \frac{\sum_{v=1}^q \omega_v}{2q}. \quad (11)$$

Eventually, using equation (7) we conclude:

$$\pi^*(u, v) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}, \forall (u, v).$$

Remark that, since Condition (2) applies, the  $\pi^*$  expressed in the previous equation are nonnegative. We will go back to this expression, in the next sections and develop a deeper focus on it, explaining the true meaning of the term "indetermination" as well as some other consequences.

## 2.4 Expected difference between coupling

Both coupling functions are extracted from an optimal transport problem concentrating values around the uniform distribution. Hence differences between them should be small in a certain sense. We provide in this section a measure of their proximity. We evaluate the expected value of a norm between the two couplings under uniform laws. More precisely we suppose that the two margins  $\mu$  and  $\nu$  are random and follow the Dirichlet's law (basically the uniformity on probability distributions). We remind here the form of that law for our application.

### Definition 3 (Dirichlet's Law)

The density of a Dirichlet law  $\mathcal{D}_p$  representing a uniform law among probability law on  $p$  elements is expressed as follows:

$$f(\mu_1, \dots, \mu_p) \prod_{k=1}^p d\mu_k = \frac{1}{B(p)} \prod_{k=1}^p \mu_k^0 \prod_{k=1}^p d\mu_k = \frac{1}{B(p)} \prod_{k=1}^p d\mu_k$$

where  $B$  is the multinomial Beta function.

Having expressed a density function for  $\mu$  and  $\nu$  (replace  $p$  by  $q$ ), we apply to them two coupling functions  $C^+$  and  $C^\times$ . As a distance, we define:

$$\Delta_p = \mathbb{E}_{(\mu, \nu) \sim \mathcal{D}_p \otimes \mathcal{D}_q} \left[ \sum_{u=1}^p \sum_{v=1}^q [(\mu \otimes \nu)_{u,v} - (\mu \oplus \nu)_{u,v}]^2 \right]$$

and compute its value through the sequence:

$$\begin{aligned} \Delta_p &= \mathbb{E}_{(\mu, \nu) \sim \mathcal{D}_p \otimes \mathcal{D}_q} \left[ \sum_{u=1}^p \sum_{v=1}^q \left[ \left( \mu_u - \frac{1}{p} \right) \left( \nu_v - \frac{1}{q} \right) \right]^2 \right] \\ &= \mathbb{E}_{\mu \sim \mathcal{D}_p} \left[ \sum_{u=1}^p \left( \mu_u - \frac{1}{p} \right)^2 \right] \mathbb{E}_{\nu \sim \mathcal{D}_q} \left[ \sum_{v=1}^q \left( \nu_v - \frac{1}{q} \right)^2 \right] \\ &= pq \mathbb{E}_{\mu \sim \mathcal{D}_p} \left[ \left( \mu_1 - \frac{1}{p} \right)^2 \right] \mathbb{E}_{\nu \sim \mathcal{D}_q} \left[ \left( \nu_1 - \frac{1}{q} \right)^2 \right] \end{aligned}$$

Now, we notice that we need to compute the variance of  $\mathcal{D}_p$ ; it holds:

**Proposition 1 (Variance of Dirichlet law)**

$$\mathbb{V}_{X \sim \mathcal{D}_p}[X] = \frac{p-1}{p^2(p+1)}$$

Proposition 1 in particular, implies that margins will concentrate their values around  $\frac{1}{p}$  and  $\frac{1}{q}$  respectively as soon as  $p$  or  $q$  increases respectively. As we notice that couplings equal each other when any margin is uniform, this should imply that  $\Delta_p$  converges to 0 if any of the two increases. This is exactly what happens, as we have:

$$\Delta_p = \frac{1}{pq} \left( \frac{p-1}{p+1} \cdot \frac{q-1}{q+1} \right) \leq \frac{1}{pq}$$

## 2.5 Structural Justification based upon an axiomatic result of Imre Csiszár

Although it seems arbitrary, our restriction to these two previous coupling functions is all but a fortuitous decision: in [12], Csiszár actually shows that, provided additional intuitive properties, we must restrict ourselves to use either least square or maximum entropy as canonic "distances" between probability distributions.

Let us rewrite our transport problems in terms with the notations of [12]. We notice that Problems 3 and 4 aim at reducing a distance from  $\pi$  to the uniform law, where  $\pi$  must satisfy constraints on its margins leading to an eligible space  $L_{\mu,\nu}$  inside the simplex  $S_D$ ,  $D > 0$ . In the first problem, the cost function is the entropy while in the second it is the  $\mathbb{L}_2$  norm.

A general question is how to adapt a "prior guess"  $u^0$  to verify a list of constraints. Let us say that  $u^0$  lives in  $S_D$  while the given constraints define a subspace  $L \in \mathcal{L}$  ( $\mathcal{L}$  is the space of subspaces of  $S_D$  spanned by a finite list of affine constraints, see [12] for more details). To formalize it, Csiszár defines a *projection rule*  $\Pi$  as a function whose input is a set  $L \in \mathcal{L}$  and which generates a method  $\Pi_L$  to project any prior guess  $u^0$  to a vector in  $L$ :

$$\begin{aligned} \Pi : \mathcal{L} &\rightarrow (S_D \rightarrow S_D) \\ L &\rightarrow \Pi_L : (u^0 \rightarrow \Pi_L(u^0) \in L) \end{aligned}$$

The article then introduces a collection of "natural" properties that we gather hereafter.

- *consistency*: if  $L' \subset L$  and  $\Pi_L(S_D) \subset L'$  then  $\Pi_{L'} = \Pi_L$ ; basically, if the result of a projection to a bigger space is always inside a smaller, then the projection on the two spaces are equivalent.
- *distinctness*: if  $L$  and  $L'$  are defined by a unique constraint and they are not equal, then  $\Pi_L \neq \Pi_{L'}$  (unless they both contains the initial prior guess). Typically, in  $\mathbb{R}^2$ , minimizing  $\|\cdot\|$  on two lines returns a different result as soon as they do not both contain 0.
- *continuity*:  $\Pi$  is continuous with regard to  $L \in \mathcal{L}$ ; it has a continuous relationship with constraints.
- *scale invariant*:  $\Pi_{\lambda L}(\lambda u) = \lambda \Pi_L(u)$  for any positive  $\lambda$  and any  $u \in S_D$ .

- *local*: for any subset  $J \subset \{1, \dots, D\}$ ,  $(\Pi_L)_J = (\Pi_{L'})_J$  as soon as  $L_J = L'_J$  where  $L_J$  means we only keep constraints dealing with coordinates in  $J$  and  $(\Pi_L)_J$  is the restriction of the resulting vector of  $\Pi_L$  to the  $J$  coordinates. This property indicates that the result of  $\Pi$  on a set of coordinates, only depends on constraints applied to these coordinates.
- *transitive*: for any  $L' \subset L$ ,  $\Pi_{L'} = \Pi'_L \circ \Pi_L$ . We can first project on a bigger space without affecting the result.

The main result of [12] states the following:

**Theorem 3 (Two canonical projections)**

*Only two projection rules respect all the conditions listed above:*

$$\begin{aligned} \Pi^2 : L &\rightarrow \Pi_L^2 : (u^0 \rightarrow \operatorname{argmin}_{v \in L} \|v - u^0\|_2) \\ \Pi^{KL} : L &\rightarrow \Pi_L^{KL} : \left( u^0 \rightarrow \operatorname{argmin}_{v \in L} \sum_{d=1}^D v_d \ln \left( \frac{v_d}{u_d^0} \right) \right) \end{aligned}$$

where  $\Pi^{KL}$  amounts to project using the "Kullback-Leibler" divergence.

To come back to our transport problem, the "prior guess" is the uniform law while the subspace  $L \subset S_D$  is defined using the margin constraints forced by  $\mu$  and  $\nu$ . Then, provided we verify the afore-mentioned properties, the two cost functions we used cover an exhaustive view. Eventually it justifies that the two graph clustering criteria comparing the neighborhood to each equilibrium (Definition 10 and Equation (15)) are canonical.

### 3 Monge properties: a justification of the $\oplus/\otimes$ notation

We introduce two classes of matrices, the first one is attributed to Gaspard Monge, from a basic idea appearing in his 1781 paper, (incidentally see [7], where a reference is given to Alan Hoffman<sup>1</sup> who first coined that point and consequently proposed the name: Monge's Matrices). For each of those Monge's matrices, we point out some remarkable equalities and, moreover, we link them to a corresponding coupling function of section 2.

#### 3.1 Monge property – "Indetermination"

To introduce Monge's properties, we follow the exhaustive work of Rainer Burkard, Bettina Klinz and Rüdiger Rudolf exposed in the 66-page-long article [7] and begin with definition 4.

---

<sup>1</sup> In 1961 Alan Hoffman (IBM Fellow and US Science Academy member) rediscovered Monges's observation see [19]. Hoffman showed that the Hitchcock–Kantorovich transportation problem can be solved by a very simple approach if its underlying cost matrix satisfies those Monge's properties

**Definition 4 (Monge and anti-Monge matrix)**

A  $p \times q$  real matrix  $c_{u,v}$  is said to be a Monge matrix if it satisfies:

$$c_{u,v} + c_{u',v'} \leq c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

and an anti-Monge matrix if:

$$c_{u,v} + c_{u',v'} \geq c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

**Remark 3 (full-Monge matrix)**

The important case for our purpose is the equality case when a matrix is both Monge and anti-Monge, we will call this situation "full-Monge" matrix.

$$c_{u,v} + c_{u',v'} = c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

Although it is poorly studied, the last introduced equality fits perfectly our purpose. The inequalities on the contrary, are common and can be met in diverse situations such as cumulative distribution functions, or copula theory.

**Remark 4 (Adjacent cells)**

A straightforward but important derived property is the local adjacency cells equality: it is sufficient (hence necessary and sufficient) to satisfy the property of the Remark 3 on adjacent cells, to ensure the obtainment of a "full-Monge" matrix behavior for the global set of cells i.e.

$$c_{u,v} + c_{u+1,v+1} = c_{u+1,v} + c_{u,v+1} \quad \forall 1 \leq u \leq p, 1 \leq v \leq q.$$

Remark 4 is a key property to study Monge matrices since it gives a direct  $\mathcal{O}(pq)$  algorithm to verify if a matrix is Monge.

Besides, a question emerges: which density function verifies the full Monge property? The following Proposition 2 solves this question.

**Proposition 2 (full-Monge matrix is equivalent to "Indetermination")**

A "full Monge matrix" exactly corresponds to an "indetermination coupling".

*Proof*

Summing on  $u'$  and  $v'$  the equality of Remark 3 we straightforwardly obtain:

$$\sum_{u'} \sum_{v'} (c_{u,v} + c_{u',v'} - c_{u',v} - c_{u,v'}) = pq c_{u,v} + c_{\cdot,\cdot} - q c_{\cdot,v} - p c_{u,\cdot} = 0$$

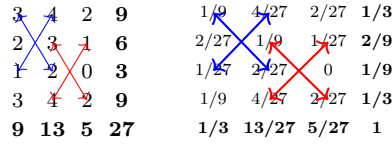
which rewrites

$$c_{u,v} = \frac{c_{u,\cdot}}{q} + \frac{c_{\cdot,v}}{p} - \frac{c_{\cdot,\cdot}}{pq}$$

By summarizing properties we get the following Theorem 4.

**Theorem 4 (full-Monge matrices)**

$\pi$  representing a probability matrix, the following properties are equivalent.



**Fig. 1** Example of an indetermination coupling (Statistical counting vs Probability forms)

1.  $\pi$  is a full-Monge matrix or equivalently all  $2 \times 2$  sub-tables  $\{u, v, u', v'\}$  extracted from  $\pi$  have the same sum on their diagonal and anti-diagonal
2.  $\pi_{u,v} = \pi_{u,v}^+ = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$
3.  $\pi$  optimizes Problem 4 for some given margins

The first property of Theorem 4 is illustrated on Figure 1 and justifies the  $\oplus$  notation assigned to "indetermination". Indeed, if we take blue and red arrows we get the same resulting value: 0. Using the contingency form:

$$\begin{aligned} \text{blue arrows} &: 3 + 2 - 1 - 4 = 0 \\ \text{red arrows} &: 3 + 2 - 4 - 1 = 0 \end{aligned}$$

Equality obviously remains true for the probability form.

### 3.2 Log-Monge property – Independence

We present hereafter a similar class of Matrices related, now, to independence: called Log-Monge matrices. They are built on the same principle as before through Definition 5.

#### Definition 5 (full-log-Monge Matrices)

A strictly positive  $p \times q$  matrix  $c_{u,v}$  is "full-log-Monge" when:

$$\ln(c_{u,v}) + \ln(c_{u',v'}) = \ln(c_{u',v}) + \ln(c_{u,v'}) \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

To immediately get the correspondence between full-Monge and full-log-Monge, we propose a transposition from a property to another using logarithm in Remark 7. It supposes matrices to be strictly positive (i.e., for our probability application, the whole discrete space must be reached).

#### Remark 5 (From Log-Monge to Monge)

We easily verify that  $c$  satisfies condition proposed in Definition 5 if and only if  $\ln(c)$  verifies the equivalent condition in Definition 4 where logarithm is taken element-wise.

Using Remark 5, we can check that full-log-Monge property is linked to "independence coupling"; without detailing their obtainment, we gather those results within Theorem 5, dual of Theorem 4.

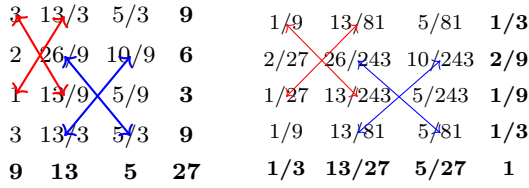


Fig. 2 Example of an "independence coupling" (Contingency vs Probability forms)

**Theorem 5 (full-log-Monge Matrices)**

$\pi$  being a strictly positive probability matrix all those properties are equivalent:

1.  $\pi_{u,v}$  is full-log-Monge or equivalently all  $2 \times 2$  sub-tables  $\{u, v, u', v'\}$  extracted from  $\pi$  have the same product on their diagonal and anti-diagonal
2.  $\pi_{u,v} = \pi_{u,v}^\times = \mu_u \nu_v$
3.  $\pi$  optimizes problem 2
- 4.

Figure 2 illustrates "full-log-Monge" matrices and their properties related to "independence"; it justifies the usual  $\otimes$  notation. Note that both those matrices (in Figure 1 and Figure 2) optimize a problem where the unique difference is the cost functions (since the margins are strictly identical).

**4 Logical "indetermination" and "Condorcet's voting equilibrium"**

In section 4, our latent goal is to better understand the "indetermination coupling", that we have until now essentially introduced from a theoretical point of view. Although obtained through a similar process, the "independence coupling" is straightforwardly linked to classical empirical experiences. The coupling defined by  $\pi^+$  does not share this latent simplicity and interpreting it, *per se*, is clearly a domain which deserves investigating. We present an attempt for helping the reader to make an accurate picture about the "indetermination" concept.

Interest for the coupling will be reinforced as we show it corresponds to the Condorcet's majority equilibrium. Defining a "for" vs "against" notion will lead us to a formal equality translating "indetermination" in another space. In fact we are faced with the famous "Condorcet's voting equilibrium", which amounts to exhibiting the situation where the number of opinions "for" balances exactly the number of opinions "against". The demonstration of this property requires the use of "Mathematical Relational Analysis" notations, which will be formally defined hereafter. In the context of this article we will not develop an exhaustive overview of this theory and its applications but merely pick up some results in connection with the goals we want to achieve; most of them are extracted from the following list of papers which gathers

some of the most important key features about the subject: [25], [22], [26], [31], [23], [1], [2].

We also interpret the equilibrium between the "yes/for" (agreements) and the "no/against" (disagreements) as a voting "indetermination situation". This implies: since the number of votes "for" equals the number of votes "against" we are in a situation, where it is impossible to take a decision. The term: "indetermination" ("indeterminacy" or "uncertainty" should have been used as well) is a formal translation of this surprising situation. First of all, let us introduce properly the Relational Analysis notations which we shall use later on.

**Definition 6 (Relational Analysis notations)**

Let  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_n)$  be two  $n$  probabilistic draws of two discrete laws  $U \sim \mu$  and  $V \sim \nu$ . We define two associated symmetric  $n \times n$  matrices  $X$  and  $Y$  by

$$X_{i,j} = \mathbf{1}_{u_i=u_j} \quad \text{and} \quad Y_{i,j} = \mathbf{1}_{v_i=v_j}, \quad \forall 1 \leq i, j \leq n$$

Basically, the two binary matrices  $X$  and  $Y$  (which correspond in fact to two binary equivalence relations based on the drawn values) represent agreements and disagreements of the two variables on a same draw of size  $n$ ; they are symmetric with 1 values on their diagonal. This relational coding has a lot of powerful properties, which can be found in the articles we mentioned beforehand.

Definition 6 immediately provides us with an algorithm to transfer contingency representations to relational ones. We would loop on  $i$  and  $j$  and fill the matrices  $X$  and  $Y$  with a 1 or a 0 according to the formula. The way back consists in noticing that:

$X_{i,j} = 1$  if and only if  $i$  and  $j$  share the same modality of  $U \sim \mu$ .

Hence we assign a modality to each class defined by the equivalence relation embedded in  $X$ : the only loss of information during this process resides in the values themselves.

Now, we are ready to present the Theorem justifying the name "indetermination":

**Theorem 6 ( $\pi^+$  and Condorcet's equilibrium)**

Let  $\pi$  be the joint probability distribution of an ordered pair of random variables  $(U, V)$  having  $p$  and  $q$  values respectively. Then  $\pi$  is an "indetermination coupling" on its margins, if and only if the expected number of "agreements" equals the number of "disagreements" on a 2 independent drawings of  $\pi$ .

*Proof*

The measure  $\pi$  is defined through its values  $\pi_{u,v}$ ,  $1 \leq u \leq p$  and  $1 \leq v \leq q$ .  $U$  and  $V$  are random variables representing its margins. Through  $n$  drawings of  $\pi$ , hence  $n$  samplings of  $(U, V)$ , two partitions of the  $n$  individuals are generated, one based on their values of  $U$ , the other on their values of  $V$ .

We will say that an agreement occurs when both partitions simultaneously gather or separate the individuals  $i$  and  $j$ . A disagreement occurs on the



contrary when a partition gathers  $i$  and  $j$  while the other one separates them. Formally, with  $X, Y$  given by Definition 6:

- $X_{i,j}Y_{i,j} = 1$ , agreement of type 11: there are  $pq$  couples of classes possible for two individuals  $i$  and  $j$  to realize this type of agreement
- $\overline{X}_{i,j}\overline{Y}_{i,j} = 1$ , agreement of type 00: there are  $p(p-1)q(q-1)$  couples of classes of this type
- $X_{i,j}\overline{Y}_{i,j} = 1$ , disagreement of type 10: there are  $pq(q-1)$  couples of classes of this type
- $\overline{X}_{i,j}Y_{i,j} = 1$ , disagreement of type 01: there are  $p(p-1)q$  couples of classes of this type

Since the number of possible couples vary we propose the following equality which establishes that the weighted number of agreements equals the weighted number of disagreements:

$$\frac{XY}{pq} + \frac{(\overline{X})(\overline{Y})}{p(p-1)q(q-1)} = \frac{X\overline{Y}}{pq(q-1)} + \frac{\overline{X}Y}{p(p-1)q} \quad (12)$$

with the scalar product notation

$$XY = \sum_{i=1}^n \sum_{j=1}^n (X_{i,j}Y_{i,j})$$

The weights appearing in Equality (12) are intrinsically important. Indeed, the just-defined equality is based on a draw of size  $n$  and precisely corresponds to a contingency indetermination. We consider two independent random draws under  $\pi$ :  $(u_i, v_i)$  and  $(u_j, v_j)$  and introduce a probabilistic equality based on our 2 draws.

$$\frac{\mathbb{E}_{\pi \otimes \pi}(X_{i,j}Y_{i,j})}{pq} + \frac{\mathbb{E}_{\pi \otimes \pi}(\overline{X}_{i,j}\overline{Y}_{i,j})}{p(p-1)q(q-1)} = \frac{\mathbb{E}_{\pi \otimes \pi}(X_{i,j}\overline{Y}_{i,j})}{pq(q-1)} + \frac{\mathbb{E}_{\pi \otimes \pi}(\overline{X}_{i,j}Y_{i,j})}{p(p-1)q} \quad (13)$$

Let us prove now that equality (13) precisely occurs when  $\pi$  equals the indetermination coupling of its margins with the formula introduced in Theorem 2.

- $\mathbb{E}_{\pi \otimes \pi}(X_{i,j}Y_{i,j}) = \sum_{u_i, v_i} \sum_{u_j, v_j} \pi_{u_i, v_i} \pi_{u_j, v_j} \mathbf{1}_{u_i=u_j \& v_i=v_j} = \sum_{u, v} \pi_{u, v}^2$
- $\mathbb{E}_{\pi \otimes \pi}(\overline{X}_{i,j}\overline{Y}_{i,j}) = \sum_{u_i, v_i} \sum_{u_j, v_j} \pi_{u_i, v_i} \pi_{u_j, v_j} \mathbf{1}_{u_i \neq u_j \& v_i \neq v_j} = \sum_{u, v} \pi_{u, v} (1 - \pi_{u, \cdot} - \pi_{\cdot, v} + \pi_{u, v})$
- $\mathbb{E}_{\pi \otimes \pi}(X_{i,j}\overline{Y}_{i,j}) = \sum_{u_i, v_i} \sum_{u_j, v_j} \pi_{u_i, v_i} \pi_{u_j, v_j} \mathbf{1}_{u_i=u_j \& v_i \neq v_j} = \sum_{u, v} \pi_{u, v} (\pi_{u, \cdot} - \pi_{u, v})$  (and similarly for  $\mathbb{E}_{\pi \otimes \pi}(\overline{X}_{i,j}Y_{i,j})$ )

Inserting this into equation (13), we get:

$$\begin{aligned} & \frac{\sum_{u,v} \pi_{u,v}^2}{pq} + \frac{\sum_{u,v} \pi_{u,v}(1 - \pi_{u,\cdot} - \pi_{\cdot,v} + \pi_{u,v})}{p(p-1)q(q-1)} \\ &= \frac{\sum_{u,v} \pi_{u,v}(\pi_{u,\cdot} - \pi_{u,v})}{pq(q-1)} + \frac{\sum_{u,v} \pi_{u,v}(\pi_{\cdot,v} - \pi_{u,v})}{p(p-1)q} \end{aligned}$$

equivalently:

$$\begin{aligned} & (p-1)(q-1) \sum_{u,v} \pi_{u,v}^2 + \sum_{u,v} \pi_{u,v}(1 - \pi_{u,\cdot} - \pi_{\cdot,v} + \pi_{u,v}) \\ &= (p-1) \sum_{u,v} \pi_{u,v}(\pi_{u,\cdot} - \pi_{u,v}) + (q-1) \sum_{u,v} \pi_{u,v}(\pi_{\cdot,v} - \pi_{u,v}) \end{aligned}$$

regrouping the similar terms yields:

$$pq \sum_{u,v} \pi_{u,v}^2 - p \sum_u \pi_{u,\cdot}^2 - q \sum_v \pi_{\cdot,v}^2 + 1 = 0$$

by developing, we verify that it corresponds to:

$$pq \sum_{u,v} (\pi_{u,v} - \pi_{u,\cdot}/q - \pi_{\cdot,v}/p + 1/pq)^2 = 0$$

Finally it holds:

$$\pi_{u,v} = \frac{\pi_{u,\cdot}}{q} + \frac{\pi_{\cdot,v}}{p} - \frac{1}{pq}$$

We have proved that  $\pi = \pi^+$  if and only if the expected number of normalized agreements equals the expected number of disagreements on a 2-sized draw.

In another article [5] we interpret the cost function linked with indetermination as a way to reduce couple matchings and show the Condorcet's equilibrium conveys useful applications as it "hides" the underlying distribution.

## 5 Application to network clustering

### 5.1 Introduction

We limit ourselves to the Louvain Algorithm, applied as a heuristic to optimize a global objective function. In a few words let us say that the global optimization is obtained iteratively by maximizing a local criterion which expresses how much two nodes are *similar*.

Conde-Céspedes, in her thesis [10], gathered a large amount of network clustering criteria, coming from the scientific literature; she took advantage of this task to give them a category label, depending on their relationship with both independence and indetermination. She compared them according to their ability to perform on various networks, and collected and stored the obtained results. Although we adopt quite the same perspective than Patricia Conde-Céspedes, we restrict ourselves to focus our study on the two canonic

deviations: "deviation to independence" and "deviation to indetermination" which we reintroduce hereafter within the network theoretical context.

Let us first give some usual definitions for a graph:

**Definition 7 (Weighted graph (network))**

A weighted graph  $G$ , is a graph which contains  $n$  nodes  $1 \leq i \leq n$ , connected to each other through edges  $(i, j)$  endowed with weights  $a_{i,j}$  (representing a weighted incidence matrix). We also introduce the total weight  $2M = \sum_{i,j} a_{i,j}$ .

A basic way to randomly generate a network is through the Gilbert's distribution:

**Definition 8 (Gilbert)**

Fixing a number  $n$  of nodes and  $\epsilon \in [0, 1]$ , we link any set of two nodes by independently drawing through a Bernoulli law with parameter  $\epsilon$  leading to a 0 – 1 weight. The obtained network is non-directed and each weight is either 0 or 1.

**Remark 6**

Adding a parameter  $p$  representing maximum weight, we can easily create a weighted graph by drawing a Binomial law with parameter  $(\epsilon, p)$  while linking ordered pairs (instead of sets) generates directed networks.

As mentioned in Section 1, our work will be devoted to the research of clusters or cliques within a given graph  $G$ . These are defined through an equivalence relation as specified in Definition 9:

**Definition 9 (Graph clustering)**

Let  $x$  be a matrix representation of a binary equivalence relation, the result of the clustering of a graph  $G$ . Then  $x_{i,j}$  equals 1 if and only if the two nodes  $i$  and  $j$  are in the same class for  $x$ , else equals 0.

Clustering algorithms aim at providing clusters maximizing internal similarities as well as minimizing external ones. A first option is to take as input the number  $K$  of classes we are looking for, together with an associated distance (or dissimilarity index) and come up with a list of best representatives or "centroids" for each class. The K-means algorithm whose idea goes back to the fifties (see<sup>2</sup> [34]) typically illustrates this option. A second option, is to construct a local criterion  $c$  which assigns a weight  $c_{i,j}$  to each ordered pair  $(i, j)$  of nodes based on their similarity; the more similar they are, the higher the criterion. We then build up a global criterion by summing up the local values  $c_{i,j}$  if and only if  $i$  and  $j$  are in the same cluster as proposed in Problem 5.

---

<sup>2</sup> factually this is the method of S. Lloyd(1957) rewritten by E.W. Forgy (1965) which corresponds to the oldest version of the K-means really used

**Problem 5 (Generic clustering problem)**

$$\max_x M(c, x) = \sum_{i=1}^n \sum_{j=1}^n c_{i,j} x_{i,j}$$

subject to:

$x$  is an equivalence relation (see Definition 6)

Let us first note that, as stated in [23], [25] and [31] an equivalence relation constraint can be written as :

- $x_{i,i} = 1, \forall 1 \leq i \leq n$  (reflexivity)
- $x_{i,j} = x_{j,i}, \forall 1 \leq i, j \leq n$  (symmetry)
- $x_{i,j} + x_{j,k} - x_{j,i} \leq 1, \forall 1 \leq i, j, k \leq n$  (transitivity)

Thanks to the linearity of these constraints, in addition to the linear expression of the criterion itself, the problem 5 although *a priori* NP-hard can be exactly solved (under unimodular matrix conditions) through the integer relaxation of a good existing 0-1 linear programming code (see [25]). But in the context of network clustering, the size  $n$  of the problem (here the number of nodes) can be really huge (millions for social networks) and the direct solving by linear programming, even specially tuned, is no longer possible; therefore, the use of robust heuristics becomes mandatory. As mentioned beforehand, the Louvain Algorithm (see [17] or [28]) belongs to this set of methods: it does not systematically provide us with an exact optimal result merely a good approximate one. We concentrate on the analysis of two canonic costs for which historical experiments are reported and explained in the light of the previous sections.

*5.1.1 Original Modularity – "Independence"*

The original and famous Newman-Girvan's presentation of a global criterion for networks clustering, see [17] or [28], has been introduced in the Louvain algorithm together with a global cost called "Modularity" defined by:

**Definition 10 (Modularity)**

Given a partition  $x_{i,j}$  and a graph  $G$  with weighting function  $a$  on its edges, the global modularity is:

$$M^\times(G, x) = \frac{1}{2M} \sum_{i,j} \left[ a_{i,j} - \frac{a_{i,\cdot} a_{\cdot,j}}{2M} \right] x_{i,j}. \quad (14)$$

Let us first remark that the original modularity  $M^\times$  is nothing but the generic global cost function defined though Problem 5 with

$$c_{i,j} = m^\times(G)_{i,j} = \frac{a_{i,j}}{2M} - \frac{a_{i,\cdot} a_{\cdot,j}}{(2M)^2}$$

and that the local gain  $m^\times(G)_{i,j}$  of putting two nodes in the same class measures a local deviation to independence. Indeed, using Definition 7, with

$\pi_{i,j} = \frac{a_{i,j}}{2M}$  as a probability measure on  $\{1 \dots n\}^2$  and margins  $\mu_i = \frac{a_{i,\cdot}}{2M}$ ,  $m^\times$  rewrites:

$$m^\times(G)_{i,j} = 2M (\pi_{i,j} - \mu_i \mu_j)$$

thus being a canonic *deviation to independence* criterion.

A second remark is that as  $m^\times(G)_{i,j}$  expression does not contain any absolute value or square elevation. Then unconnected nodes will lead to negative weights preventing them from being assigned to the same class. If they are connected, the importance of  $m^\times(G)_{i,j}$  evolves positively as  $i$  and  $j$  have less edges ( $a_{i,\cdot}$  and  $a_{\cdot,j}$  small); here again this implies an appropriate behavior. More precisely, since independence ensures as uniform a coupling as possible with fixed margins (this is a solution of Problem 2),  $m^\times$  appears as a fair construction. The criterion basically measures a distance between the observed linkage weight and an expected flat weight given by the average neighborhood.

### 5.1.2 Extended Modularity – "Indetermination"

We suggest an expression  $m^+(G)_{i,j}$  which represents a deviation to indetermination. It will be used as a local cost function in Problem 5 leading to a slightly different global formula  $M^+(G, x)$  to optimize locally

$$m^+(G)_{i,j} = a_{i,j} - \frac{a_{i,\cdot}}{n} - \frac{a_{\cdot,j}}{n} + \frac{2M}{n^2}$$

Just as  $m^\times$ , it rewrites as a canonic *deviation to indetermination*

$$m^+(G)_{i,j} = 2M * \left( \pi_{i,j} - \frac{\mu_i}{n} - \frac{\mu_j}{n} + \frac{1}{n^2} \right)$$

The global criterion follows:

$$M^+(G, x) = \sum_{i,j} \left[ a_{i,j} - \frac{a_{i,\cdot}}{n} - \frac{a_{\cdot,j}}{n} + \frac{2M}{n^2} \right] x_{i,j} \quad (15)$$

We have seen that both couplings share a lot of properties shown in Section 3 and Section 4. In the same way, Patricia Conde-Céspedes noticed that a lot of statistical criteria measuring variables correlation are based either on a "distance to independence", or on a "distance to indetermination" (see [10]). According to these remarks, our canonical *deviation to indetermination* criterion  $M^+$  deserves the same types of use as those dedicated to the Newman Girvan's  $M^\times$ .

## 5.2 Summary of an application to various networks

Now that the two criteria have been theoretically grounded on their canonical structure, we gather in Table 1 the number of clusters found by Patricia Conde-Céspedes, who applied both on the same empirical networks. She got similar results, as expected beforehand on a bench of experimental graphs she properly

defines in [10] and for which we gather her results in Table 1 which can be read as follows: for example, the "Internet" network contains 69,949 nodes with 351,280 edges; if we apply the Louvain algorithm to it, with the global criteria  $M^\times$  we usually find 46 communities, while  $M^+$  leads to 39. As anticipated in Section 2.4 the criteria are (in average) very close (see for instance the Amazon case); consequently their resulting effect on various networks is quite similar. Section 5.3 of the present paper provides the reader with an explanation of the assertion Patricia Conde-Céspedes experimented. However, graphs in the real world may greatly differ from the simplified Gilbert's model presented below (see for instance the YouTube example of the table).

**Table 1** Number of classes found by each criteria on various networks

	Karate	Football	Jazz	Internet	Amazon	YouTube
N (nb nodes)	34	115	198	69 949	334 863	1 134 890
M (sum of weights)	78	613	2 742	351 280	925 872	2 987 624
Number of classes for criteria $M^\times$	4	10	4	46	250	5 567
Number of classes for criteria $M^+$	4	10	6	39	246	13 985

### 5.3 Gilbert Experimental Tests

As already mentioned, Solving problem 5 is NP-hard so that we cannot expect precise results, neither about the number of classes for a given criterion, nor about their composition. Nevertheless, we can compare directly local values of the criterion to extrapolate a common or a distinct global run when iteratively optimized.

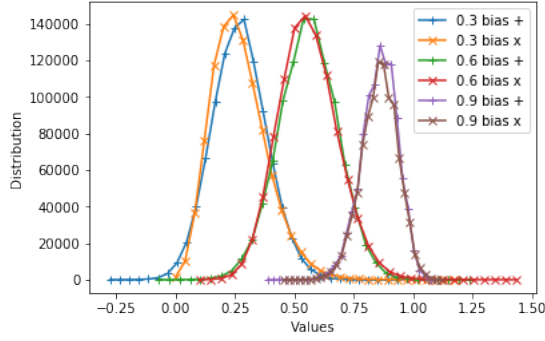
We propose a comparative trial based on Gilbert's networks to spot differences or similarities between  $m^\times(G)_{i,j}$  and  $m^+(G)_{i,j}$  values. The aim is to observe the distribution of both criteria on a typical network. First, to simplify observations and as only the reference cost (*i.e.* an equilibrium of Section 2) varies between  $m^+$  and  $m^\times$ . It corresponds to the criteria minus  $a_{i,j}$ ; it is formally defined in Definition 11. Then, we generate 1,000,000 networks randomly, compute each criterion on a random pair of nodes and store the reference cost; the results are displayed on Figure 3.

**Definition 11 (Bias or reference cost)**

*The two bias derived from  $m^\times$  and  $m^+$  are respectively:*

$$b_{i,j}^\times = \frac{a_{i,\cdot} a_{\cdot,j}}{2M} \quad \text{and} \quad b_{i,j}^+ = \frac{a_{i,\cdot}}{n} + \frac{a_{\cdot,j}}{n} - \frac{2M}{n^2}$$

On Figure 3 we observe that the distributions of both biases are similar for any values of  $\epsilon$ . Indeed, the curves are identical on their core values (those with a number of realizations over 20,000). It illustrates their common origin



**Fig. 3** Empirical distribution of the two reference costs  $b_{i,j}^+$  and  $b_{i,j}^x$  for  $\epsilon$  in  $[0.3, 0.6, 0.9]$ ; X-axis gives the values of the bias, Y-axis gives the corresponding number of realizations

which amounts to flatten a distribution (Section 2) and leads to a small expected difference (Section 2.4). Their common mean is equal to  $\epsilon$  as can be easily derived from the formulas. A difference nevertheless remains on extreme values particularly visible for  $\epsilon = 0.3$ . Let us now theoretically compute both distributions under Gilbert's networks to confirm their symmetry.

### Proposition 3 (Probability values)

Let  $b$  be a binary value,  $b \leq n_i \leq n$  and  $b \leq n_j \leq n$ , then:

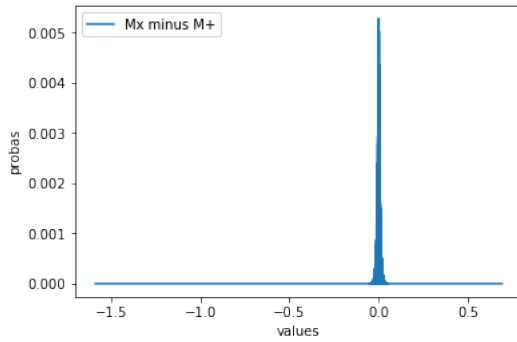
$$\begin{aligned} & \mathbb{P}(a_{i,j} = b, a_{i,\cdot} = n_i, a_{\cdot,i} = n_j) \\ &= \mathbb{P}(a_{i,j} = b) \times \mathbb{P}(a_{i,\cdot} = n_i | a_{i,j} = b) \times \mathbb{P}(a_{\cdot,i} = n_j | a_{i,j} = b) \\ &= \epsilon^b (1-\epsilon)^{1-b} \binom{n-1}{n_i-b} \epsilon^{n_i-b} (1-\epsilon)^{n-1-n_i+b} \binom{n-1}{n_j-b} \epsilon^{n_j-b} (1-\epsilon)^{n-1-n_j+b} \end{aligned}$$

The corresponding value  $m_{i,j}^+$  and  $m_{i,j}^x$  associated with a group  $(b, n_i, n_j)$  of the parameters being evident, we propose Figure 4 which represents the difference between the theoretical distributions of the two criteria with  $\epsilon = 0.3$ .  $b^x$  and  $b^+$  have distinct forms but their proximity on highly probable values, given on Figure 4, illustrates Section 2.4: if we couple two variables with  $n$  margins, the expected difference is less than  $\frac{1}{n^2}$ .

Extreme values, on the contrary may differ drastically. Though it seems the opposite on Figure 3 as  $m^+$  comes with higher values than  $m^x$ , it is consistent because of the *minus* sign in the formula linking  $m$  with  $b$ . Having noticed that  $b^+$  and  $b^x$  differ on their extreme values, we compute them on a general Gilbert network (respecting the common value of  $2M = n^2\epsilon$ ), and obtain the bounds:

$$-\epsilon \leq b^+ \leq \frac{n}{n} + \frac{n}{n} - \epsilon = 2 - \epsilon \quad \text{and} \quad 0 \leq b^x \leq \frac{n \times n}{n^2\epsilon} = \frac{1}{\epsilon} \quad (16)$$

As expected from Figure 4 the difference between extreme values is arbitrarily high. It eventually shows that on average, both canonical criteria will



**Fig. 4** Theoretical distribution of the difference  $m^\times(G)_{i,j} - m^+(G)_{i,j}$  (same as  $b^+_{i,j} - b^\times_{i,j}$ ) on generated graphs

share a similar behavior hence shall be applied indifferently when the network's neighborhood weighting  $a$  is close to uniform. Precisely, in any real application the chosen criterion detects, on each iteration, a link among the neighborhood, which is unexpected with respect to the equilibrium (either independence or indetermination). The difference between the two equilibria is so small that if one detects a link, the other will too, so that we expect the resulting classes to be comparable if not identical. Indeed, we know we can build, networks on which  $\mathcal{M}^+$  and  $\mathcal{M}^\times$  deliver completely different results as an application of Equation (16).

## 6 Conclusions

First, we followed the historical line and introduced two bases from Discrete Optimal Transport Theory: *independence* and *indetermination*. As recalled, the first one is the most intuitive and frequently used in mathematical articles as well as experimented in real life. The second notion appeared more surprising, poorly studied in the statistical literature but more commonly used by people working on Mathematical Relational Analysis, Voting Theory, and Analysis of Variance. Together, they cover the only two canonic projection costs as quoted in [12].

To illustrate the usefulness of the parallel construction, we turned to applications and completed the track followed by Patricia Conde-Céspedes in her thesis [10]. She gathered a list of network clustering criteria and classified them according to their deviation to one of the mentioned coupling functions. Section 5 reports a further analysis of the two canonical criteria. It gathers results about the general similarity of their applications to various networks as well as their extreme values to highlight their differences.

In each section, from optimal transport to networks, we insisted on the parallel between the two notions together with their differences. As quoted



beforehand, they appear as the two unique canonic structural solutions. Generally, the differences between them need scanning up, either to coin a macro criteria, or to chose wisely between one and the other depending on the structure of the network. In any case, the traditional use of *independence* at the expense of *indetermination* needs to be further investigated and explained.

**Acknowledgements** We thank the editor and two anonymous referees for their valuable comments.

## References

1. Ah-Pine, J.: Sur des aspects algébriques et combinatoires de l'analyse relationnelle: applications en classification automatique, en théorie du choix social et en théorie des tresses. Ph.D. thesis, Paris 6 (2007)
2. Ah-Pine, J.: On aggregating binary relations using 0-1 integer linear programming. In: ISAIM, pp. 1–10 (2010)
3. Asano, T., Bhattacharya, B., Keil, M., Yao, F.: Clustering algorithms based on minimum and maximum spanning trees. In: Proceedings of the fourth annual symposium on Computational Geometry, pp. 252–257 (1988)
4. Bertrand, P.: Transport optimal, matrices de monge et pont relationnel. Ph.D. thesis, Paris 6 (2021)
5. Bertrand, P., Broniatowski, M., Marcotorchino, J.F.: Logical indetermination coupling: a method to minimize drawing matches and its applications (2020). URL <https://hal.archives-ouvertes.fr/hal-03086553>. Working paper or preprint
6. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* **2008**(10) (2008)
7. Burkard, R.E., Klinz, B., Rudolf, R.: Perspectives of Monge properties in optimization. *Discrete Applied Mathematics* **70**, 95–161 (1996)
8. Campigotto, R., Conde-Céspedes, P., Guillaume, J.L.: A generalized and adaptive method for community detection. *arXiv preprint arXiv:1406.2518* (2014)
9. Chen, M., Nguyen, T., Szymanski, B.K.: A new metric for quality of network community structure. *arXiv preprint arXiv:1507.04308* (2015)
10. Conde-Céspedes, P.: Modélisations et extensions du formalisme de l'analyse relationnelle mathématique à la modularisation des grands graphes. Ph.D. thesis, Paris 6 (2013)
11. Coron, J.L.: Quelques exemples de jeux à champ moyen. Ph.D. thesis, Paris Sciences et Lettres (ComUE) (2017)
12. Csiszár, I., et al.: Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems. *The annals of statistics* **19**(4), 2032–2066 (1991)
13. Doreian, P., Batagelj, V., Ferligoj, A.: *Advances in Network Clustering and Blockmodelling*. John Wiley & Sons (2020)
14. Fortunato, S.: Community detection in graphs. *Physics reports* **486**(3-5), 75–174 (2010)
15. Fortunato, S., Barthelemy, M.: Resolution limit in community detection. *Proceedings of the national academy of sciences* **104**(1), 36–41 (2007)
16. Fréchet, M.: Sur les tableaux de corrélations dont les marges sont données. *Annales de l'Université de Lyon, Section. A* **14**, 53–77 (1951)
17. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* p. 7821–7826 (2002)
18. Hamilton, R.S., et al.: Three-manifolds with positive ricci curvature. *J. Differential geom* **17**(2), 255–306 (1982)
19. Hoffman, A.J.: On simple linear programming problems. *Proceedings of the Seventh Symposium in Pure Mathematics of the AMS* pp. 317–327 (1963)
20. Katz, L.: A new status index derived from sociometric analysis. *Psychometrika* **18**(1), 39–43 (1953)

21. Lancichinetti, A., Fortunato, S.: Limits of modularity maximization in community detection. *Physical review E* **84**(6), 066122 (2011)
22. Marcotorchino, J.F.: Utilisation des comparaisons par paires en statistique des contingences. Publication du Centre Scientifique IBM de Paris et Cahiers du Séminaire Analyse des Données et Processus Stochastiques Université Libre de Bruxelles pp. 1–57 (1984)
23. Marcotorchino, J.F.: Maximal association theory as a tool of research. Classification as a tool of research , W.Gaul and M. Schader editors, North Holland Amsterdam (1986)
24. Marcotorchino, J.F., Conde-Céspedes, P.: Optimal transport and minimal trade problem, impacts on relational metrics and applications to large graphs and networks modularity. In: International Conference on Geometric Science of Information, pp. 169–179. Springer (2013)
25. Marcotorchino, J.F., Michaud, P.: Optimisation en Analyse Ordinale des Données. Ed Masson, Paris (1979)
26. Messatfa, H.: Maximal association for the sum of squares of a contingency table. *Revue RAIRO, Recherche Opérationnelle* **24**, 29–47 (1990)
27. Nascimento, M.C.: Community detection in networks via a spectral heuristic based on the clustering coefficient. *Discrete Applied Mathematics* **176**, 89–99 (2014)
28. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. *Physical review E* **69**(2), 026113 (2004)
29. Ni, C.C., Lin, Y.Y., Luo, F., Gao, J.: Community detection on networks with Ricci flow. *Scientific reports* **9**(1), 1–12 (2019)
30. Ollivier, Y.: Ricci curvature of markov chains on metric spaces. *Journal of Functional Analysis* **256**(3), 810–864 (2009)
31. Opitz, O., Paul, H.: Aggregation of ordinal judgements based on condorcet’s majority rule. *Data Analysis and Decision Support. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg (2005)
32. Reichardt, J., Bornholdt, S.: Statistical mechanics of community detection. *Physical review E* **74**(1), 016110 (2006)
33. Sklar, A.: Random variables, joint distribution functions, and copulas. *Kybernetika* **9**(6), 449–460 (1973)
34. Steinhaus, H.: Sur la division des corps matériels en parties. *Bulletin de l’académie polonaise des sciences*, v. 4, no. 12 p. 801–804 (1957)
35. Stemmelen, E.: Tableaux d’échanges, description et prévision. *Cahiers du Bureau Universitaire de Recherche Opérationnelle* **28** (1977)
36. Wilson, A.G.: A statistical theory of spatial distribution models. *Transportation Research* **1**, 253–269 (1967)
37. Wilson, A.G.: The use of entropy maximising models. *Journal of transport economics and policy* **3**, 108–126 (1969)