A network-based method to detect patterns of local crop biodiversity: validation at the species and infra-species levels

Mathieu Thomas[1,2,3,*], Nicolas Verzelen[4], Pierre Barbillon[5], Oliver T. Coomes[6], Sophie Caillon[2], Doyle McKey[2,7], Marianne Elias[8], Eric Garine[9], Christine Raimond[10], Edmond Dounias[2], Devra Jarvis[11,12], Jean Wencélius[9], Christian Leclerc[13], Vanesse Labeyrie[13], Pham Hung Cuong[14], Nguyen Thi Ngoc Hue[14], Bhuwon Sthapit[15], Ram Bahadur Rana[16], Adeline Barnaud[17,18,19], Chloé Violon[9], Luis Manuel Arias Reyes[20], Luis Latournerie Moreno[21], Paola De Santis[22], François Massol[23]

[1] INRA, UMR 0320 / UMR 8120 Génétique Quantitative et Évolution - Le Moulon, F-91190 Gif-sur-Yvette, France

[2] CEFE UMR 5175, CNRS - Université de Montpellier - Université Paul-Valéry Montpellier - EPHE – 1919 route de Mende, 34293 Montpellier cedex 5, France

[3] CESAB/FRB, F-13857 Aix-en-Provence, France

[4] UMR 729 Mathématiques, Informatique et STatistique pour l'Environnement et l'Agronomie, INRA SUPAGRO, F-34000 Montpellier, France

[5] AgroParisTech / UMR INRA MIA, F-75005 Paris, France

[6] Department of Geography, Burnside Hall, rm 705, McGill University, 805 Sherbrooke Street West, Montreal, QC Canada H3A 0B9

[7] Institut Universitaire de France, France

[8] Institut de Systématique, Évolution, Biodiversité ISYEB - UMR 7205 – CNRS, MNHN, UPMC, EPHE Muséum national d'Histoire naturelle, Sorbonne Universités, 57 rue Cuvier, CP50, F-75005 Paris, France

[9] Université Paris Ouest / CNRS, UMR 7186 LESC, F-92000 Nanterre, France

[10] CNRS- UMR 8586 Prodig, 2 rue Valette 75005 Paris, France

[11] Bioversity International, Via dei Tre Denari 472/a, 00057 Maccarese, Rome, Italy

[12] Department of Crop and Soil Sciences, Washington State University, Pullman WA USA

[13] CIRAD, UMR AGAP, F-34398 Montpellier, France

[14] Plant Resources Center - VAAS, MARD, New Town, Southern An Khanh com., Hoai Duc dist., Hanoi, Vietnam

[15] Bioversity International, Office of Nepal, 93.4 Dharahara Marg, Fulbari, Ward # 011, Pokhara City, Nepal

[16] Local Initiatives for Biodiversity, Research and Development (LI-BIRD), P.O. Box 324, Pokhara, Kaski, Nepal

[17] IRD, UMR DIADE, F-34398 Montpellier, France

[18] LMI LAPSE, Dakar, Sénégal

[19] ISRA, LNRPV, Centre de Bel Air, Dakar, Sénégal

[20] CINVESTAV-IPN Unidad Mérida, Merida, Yucatan, Mexico

[21] Instituto Tecnológico de Conkal, Division de Estudios de Posgrdo e Investigacion Conkal, Yucatan, Mexico

[22] Bioversity International, Via dei Tre Denari 472/a, 00057 Maccarese, Rome, Italy

[23] CNRS, Unité Evolution, Ecologie et Paléontologie (EEP), CNRS UMR 8198, Université Lille, Lille, France

* Corresponding author

E-mail addresses: thomas@moulon.inra.fr

Key words: agrobiodiversity; bipartite networks; seed exchange.

Short title: Patterns of local crop biodiversity

Data availability: Code is available at: http://netseed.cesab.org/.

Abstract

In this paper we develop new indicators and statistical tests to characterize patterns of crop diversity at local scales. Households growing a large number of species or landraces are known to contribute an important share of local available diversity of both rare and common plants but the role of households with low diversity remain little understood: do they grow only common varieties – following a nestedness pattern typical of mutualistic networks in ecology – or do 'diversity poor' households also grow rare varieties? This question is pivotal in ongoing efforts to assess the contribution of small farmers to global agrobiodiversity at local scales. We develop new network-based approaches to characterize the distribution of local crop diversity at the village level (species and infra-species) and validate these approaches using meta-data sets from 10 countries. Our results highlight the sources of heterogeneity in the local crop diversity. We often identify two or more groups of households based on their different levels of diversity. In some datasets, 'diversity poor' households significantly contribute to the local crop diversity. Generally, we find that the distribution of crop diversity is more heterogeneous at the species than at the infra-species level. This analysis reveals the absence of a general pattern of crop diversity distribution independent of agro-ecological and socio-cultural context.

Contents

1 Glossary

Network: is a finite set of nodes (vertices) connected by links (edges).

Node: is a synonym of a vertex and is the fundamental unit of which graphs are formed.

Edge: an edge is a link between two vertices,every edge has two endpoints in the sets of vertices. In the particular case of bipartite networks, the two endpoints belong to two disjoint subsets of nodes, *e.g.* farmer households (H) and crops (S, species or landraces). The presence of an edge indicates that the household grows the considered crop.

Degree: the number of edges incident to a vertex, *e.g.* a household's degree is the number of crops grown by the considered household.

Interaction network: a network of nodes that are connected by features, *e.g.* in a crop-household interaction network, crops are grown by farmers that are members of the household.

Bipartite network: network whose vertices can be partitioned into two disjoint subsets (*e.g.* F to represent the farmer households and S to represent the species/landraces) such that no edge connects two vertices from F or two vertices from S.

Incidence matrix: 0/1 matrix A. Its rows are indexed by the set of households F and its columns are indexed by the set of plants S. The entry A_{ij} equals one if and only if farmer i grows plant j (see Section 4.1).

Nestedness: this index quantifies the extent to which nodes of one subset (*e.g.*: F) with low degrees are linked to nodes of the other sub-set (*e.g.*: S) with high degrees. In the example of household-plant network, it measures to what extent 'diversity poor' households grow a subset of plants grown by 'diversity rich' households.

Erdős-Rényi model: a random graph model where all the edges are drawn independently with the same probability p.

Latent block models: random graph models assuming a mixture distribution both on rows (households) and columns (plants). Households and plants are assumed to belong to blocks which are

latent (not observed). Thanks to a statistical inference procedure this block structure can be uncovered (see Section 4.3).

Configuration model: a random graph model with a prescribed degree sequence. All graphs with this degree sequence obtained by permutation are equiprobable in this model (for details see section 4.4.1).

2 Introduction

Agriculture relies on the use of crop plant species to provision human societies with food, clothing, medicinal, narcotic, fodder purpose, and building materials. Crop species have been domesticated from wild ancestors, which often display variability in traits related to their local adaptation to the environment. During domestication, only a subset of diversity from the wild ancestors was selected, and shaped in divergent ways depending on the goals of farmers, to produce a diversity of landraces, named and managed as distinct entities (Diamond, 2002). Furthermore, different crop species play distinctive, often complementary, roles in agriculture. In traditional agro-ecosystems, the end result of these processes of selection among wild diversity, divergent selection in farmers' fields and adoption of numerous kinds of crops, is a substantial diversity of cultivated plants, both in terms of the number of species and landrace diversity within species (Jarvis et al., 2008).

A primary requisite to understanding and predicting the sustainability of agricultural systems facing environmental, political, social and economical changes is to assess how these systems can maintain crop diversity (*e.g.* Samberg et al., 2013). For instance, in the case of manioc managed by Makushi Amerindians of Guyana, some varieties are specially grown for special dishes, others for another use; some grow quickly, thereby ensuring early yield, while others grow slowly and act ever-present insurance (Elias et al., 2000). Often, diversity is just valued for its own sake (Boster, 1985), or as a means to foster social relations (Heckler and Zent, 2008). Another example is the great diversity of landraces present in milpas of Yucatan, the end product of several thousand years of directed selection on maize, beans, squash and chile peppers by the region's farmers. Understanding landraces relationships, it is possible to gain insight into the cultural history of crops in Yucatan. The particular traits exhibited by local varieties grown in milpas today reflect Yucate-

can farmers' short- and long-term responses to agroenvironmental conditions, the ecological demands of crop production, and the aesthetic, culinary, and religious sensibilities of farmers (Tuxill et al., 2010). Maintaining crop diversity is of paramount importance in helping crops and farmers adapt to global changes, notably climate change (Vigouroux et al., 2011) and the increasingly rapid emergence of agricultural pests (Diamond, 2002). In the face of such change, drastically reduced diversity of crop species and varieties would inevitably lead to increasingly unpredictable yields. In addition, cultivating diverse crops and varieties at the landscape level favors ecological and economic sustainability by reducing the need for chemical inputs (Bianchi et al., 2006; Crowder et al., 2010).

From a purely biological perspective, the spatial distribution of crop diversity is expected to be partially explained by environmental factors due to the differential adaptation of crops to local conditions (Mariac et al., 2011). For instance, dry and wet climates do not require the same physiological adaptations of plants, and different limiting factors impose different strategies to cope with them. Selective pressures in cultivated environments differ from those in wild environments. However, unless massive inputs (unsustainably) free crops from environmental constraints, adaptation to local abiotic environments is expected to shape crop diversity — as it shapes the diversity of wild plants — at more or less large spatial scales, *e.g.* over latitudinal or altitudinal gradients. At fine spatial scales, local adaptation is also expected to play a role in the distribution of crop diversity, *e.g.* due to the heterogeneity of soil quality of agricultural fields and to variability in local rainfall (Fraser et al., 2012).

In addition to environmental factors, it has been argued that crop diversity can only be understood by accounting for social and cultural aspects of their contextual environment (Leclerc and Coppens d'Eeckenbrugge, 2012; Rival and McKey, 2008). Agricultural societies have shaped the diversity of their cultivated crops in ways that fitted their traditions, habits, myths, social organizations, and livelihoods (Dêletre et al., 2011; Leclerc and Coppens d'Eeckenbrugge, 2012). In fact, crops and humans have likely evolved together, as cultural practices may have been shaped by available edible plants as much as agricultural selection may have answered cultural needs. The study of crop genetic and interspecific diversity in the context of both environment- and society-driven selective pressures is now taken into account through the GxExS framework (Leclerc and

Coppens d'Eeckenbrugge, 2012).

Thus, studying the distribution of crop diversity and linking it with both social and environmental factors cannot be based on a uniquely biological perspective. However, interdisciplinary studies of the distribution of crop diversity must retain quantitative rigor and thus be based on a sound statistical framework. So far, the distribution of crop diversity has been assessed mostly through the use of diversity indices adopted from ecology and economics, *e.g.* indices of richness, evenness, concentration, etc. (textite.g. Jarvis et al., 2008). However, such indices only make use of crop diversity data as an instance of "type in location" data, and this limits the types of questions they address. For example, these indices can help explain why crops are found in the fields they are in, but not why farmers happen to cultivate this or that crop. A significant shortcoming of studies of the distribution of crop diversity is that they have failed to utilize the network* [1] feature of crops-by-farmers datasets which include social aspects such as farmer-to-farmer circulation of seeds (and other propagules) of varieties and crop species.

Our main goal in this paper is to answer the question "which households contribute, and how, to the diversity of crops grown in a given village?" by examining on inventories of crops species and landraces grown at the household level. To do so, we offer a novel methodological framework using network-based and null model-based statistical tests. From a methodological perspective, inventory datasets can be construed as bipartite networks*, namely crop-by-household interaction networks, in the same way as plant-pollinator or host-parasite interaction networks in ecology. In social network analysis, network approaches have been used to assess the properties of network processes linked to social institutions such as friendship, advice or seed exchange networks ("who interacts with whom" or "who gives to whom") (Wasserman and Faust, 1994; Lazega et al., 2012; Reyes-García et al., 2013). In ecology, on the other hand, networks have been used to study both contact networks (metapopulations or metacommunities) and structured interaction networks* such as food webs (*e.g.* host-plant networks) or mutualistic networks (*e.g.* plant-pollinator networks). When interaction partners can be clearly categorized (*e.g.* plants *vs.* pollinators; plants, herbivores and parasitoids), the use of bi- or multi-partite networks is an appropriate approach. In the present study, we develop a framework for the study of crop-by-household datasets that makes

[1] * indicates these words or expression are defined in the Glossary section

use of the bipartite nature of the data to reveal potential patterns of diversity structure at the scale of the village or of clusters of interacting villages.

Our paper offers an alternative to the nestedness* approach, for several reasons as detailed below. The study of bipartite networks in ecology is a recent endeavor (Jordano, 1987). In the last three decades, the topological properties of bipartite networks have been studied to answer a variety of questions, *e.g.* whether such networks are stable, robust to species extinctions or additions, functionally redundant, etc. (Jordano et al., 2003; Thébault and Fontaine, 2010). In particular, the nestedness of mutualistic bipartite networks often has been investigated, and studies suggest how it nestedness may be the key property explaining the dynamics and structural stability of mutualistic networks (Thébault and Fontaine, 2010). Such patterns are often explained as resulting from source-sink processes wherein species-rich locations function as sources producing many emigrating individuals which, in turn, contribute to the diversity in species-poor, sink locations Mat: [ref] (), or from feasibility constraints on the existence of specialist-specialist interactions in mutualistic networks Mat: [ref] (). In systems involving social as well as ecological processes, such as in the present case of crop-by-household interactions, one may ask whether the plants present in less diverse farms systematically comprise a subset of those cultivated in more diverse farms. Among the Duupa in northern Cameron, for example, older farmers accumulate crop diversity during over their life (sources) and become sources of diversity for young farmers (sinks) (Alvarez et al., 2005). When crops are actively cultivated by farmers, for example as staple food, copying other farmers' portfolios of crops might result in strong similarities in cultivated diversity among fields, but not necessarily following a nested pattern. Therefore, contrary to the case for ecological systems, certain mechanistic reasons may justify considering crop-by-household interactions as systematically nested, precluding explanations solely based on source-sink processes.

From a purely methodological perspective, available indices of network nestedness are quite inconsistent, both in the value of nestedness metrics and in their associated p-value when confronted with the configuration model, a null model of partner interactions constrained by degree, *i.e.* fixing the degree of rows and columns (Podani and Schmera, 2012). Therefore, nestedness is still a more or less verbal concept, its mathematical definition is in need of refinement, researchers have yet to study possible nestedness patterns in crop diversity research.

In the first section of our paper, we introduce a meta-dataset of specific and infra-specific crop diversity at the local scale in different agricultural contexts. In the second section, we describe our methodological framework, and the tests proposed, illustrated with a few toy examples, *i.e.: a hypothetical example*: (*i*) to test whether the variability in the number of connection per household and per crop type is different from random expectations under an homogeneous random graph model (Erdős-Rényi model*); (*ii*) to reveal structure (*e.g.* modules, cores, etc.) in the dataset using latent block models* (LBMs); (*iii*) to uncover "outliers" (*i.e.* farmers or crop types that do not conform to the general connection pattern) using principal component analyses (PCAs); and, (*iv*) to measure and to test the originality of farmers' contributions to overall crop diversity using beta-diversity indices. In the third section, we perform a meta-analysis applying the methodological framework to our meta-dataset, which allows us to highlight both regularities and particularities among the datasets. Overall, our approach yields graphical representations of the different tests (*e.g.* re-ordering of interactions in the case of LBMs or principal plane representations for PCAs) and non-parametric tests of our hypotheses, the significance of which is assessed through comparison with a permutation-based null model (the configuration model for graphs with given degrees). These graphical and statistical approaches are to be easily transferable to similar problems arising in other research fields, *e.g.* in ecology. Before concluding, we dedicate the final section to the discussion of the results and of the value and the limits of this approach.

3 Description of the datasets used in the meta-analysis

Fifty published or unpublished datasets dealing with crop inventories were provided by ethnobiologists, geographers, and ecologists (Table 1 and 2). These data were collected in 10 different countries (Figure 1) between 1998 and 2013. For each dataset, a partial set or the full set of households from the same village was characterized for one of the two classes of Operational Taxonomic Units (OTU) considered: the species or the infra-species level. This information was gathered through direct interviews with the cultivators of the household, a subset of them or only with the head of the household. Datasets were selected when the number of characterized households and OTU was higher than 10. For 18 datasets, information was collected at the species level (Table 1),

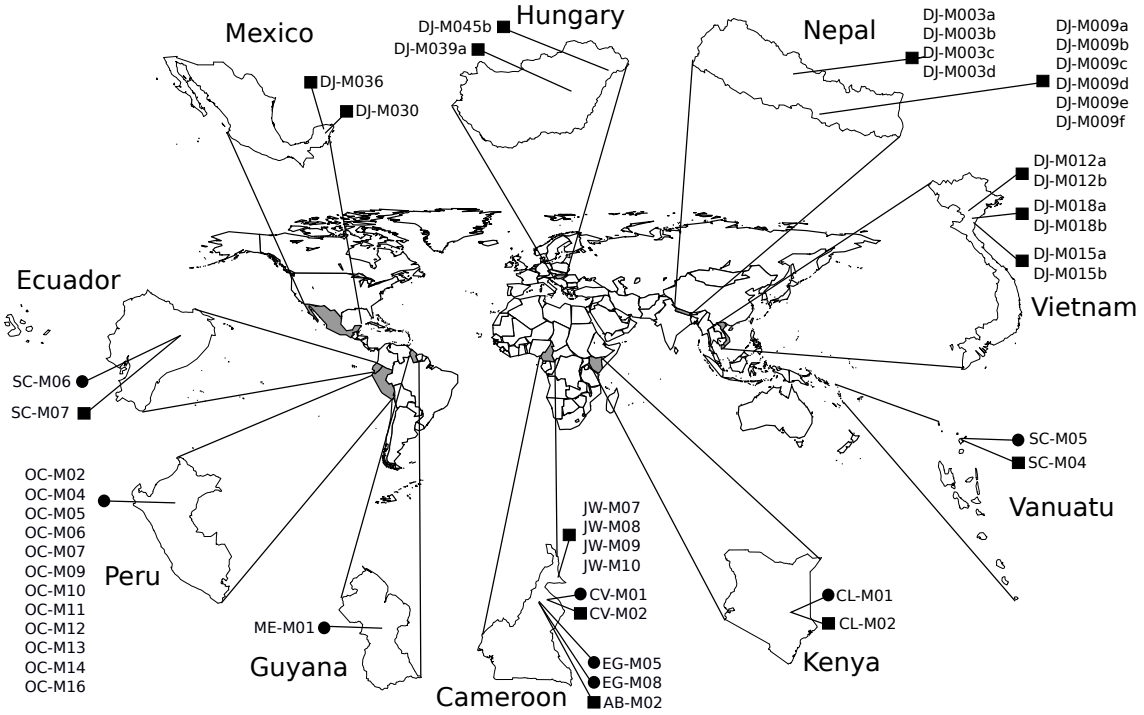

Figure 1: Map locating the different data sets used in the meta-analysis. Filled circles correspond to the data sets collected at the specific level and filled squares correspond to the data set collected at the infra-specific level

for 32 datasets, information was collected at the landrace level, which corresponds to the terminal taxon in the farmer local naming systems, covering seven different species (maize, rice, wheat, bean, manioc, taro and sorghum) which correspond to the major crops of the under area (Table 2). These species are characterized by their predominant propagation mode with partially outcrossing, outcrossing, inbreeding and clonal following the classification proposed by Jarvis et al. (2008). Data were structured following a rectangular incidence matrix* with households in rows and species or landraces in columns, and represented as a bipartite network. Data collected at the species or infra-species level represent two levels of local crop biodiversity. Underlying processes shaping the distribution of local crop diversity are assumed to be different for these two levels. Therefore, species and infra-species data are analyzed and described separately.

4 Description of the methodological framework

This section introduces the statistical framework for analyzing household-plant network data. After defining the main concepts, we detail the four main steps of the analysis. First, the degree distribution of the data is evaluated as a way to test whether a completely random model (Erdős-Renyii model) fits well the data. Second, we use a latent block model to investigate more thoroughly the structure of the network. Intuitively, this method pinpoints groups of households and groups of plants that tend to be highly connected. Then, it is tested whether this high-level structure (blocks) is not simply a consequence of low-level structures such as degree heterogeneity. These methods provide new graphical representations of the data emphasizing the studied patterns. Finally, complementary analyses based diversity measures on diversity measure are introduced. In each subsection, toy-examples illustrate the purpose, the benefits and the downsides of the proposed methods.

4.1 Mathematical formalism

In the following, we denote n the number of households, m the number of plants. The incidence matrix (with households as rows and plants as columns) that summarizes the data is noted $\mathbf{X}$, so that $X_{ij} = 1$ when household i cultivates plant j. Using this representation (see Figure 2), we can readily apply statistical methods for binary matrices.

Any incidence matrix $\mathbf{X}$ can also be treated as the adjacency matrix of some bipartite graph $\mathcal{G}$. More specifically, consider a collection of nodes corresponding to all households and all species (or landraces) and put an edge between the household i and the plant j if and only if $X_{ij} = 1$. The obtained network is bipartite (see Figure 2) as no two households and no two species are connected in the network. Building on this equivalence between incidence matrices and bipartite graphs, we can borrow methodologies developed in the field of network analysis Kolaczyk (2009).

As these two representations are equivalent, any statistical analysis could be defined either in terms of the incidence matrix $\mathbf{X}$ or in terms of the bipartite network $\mathcal{G}$. To ease the reading, this paper makes use of the incidence matrix terminology but we sometimes borrow network notations to emphasize the connection with the extant literature on network analysis.

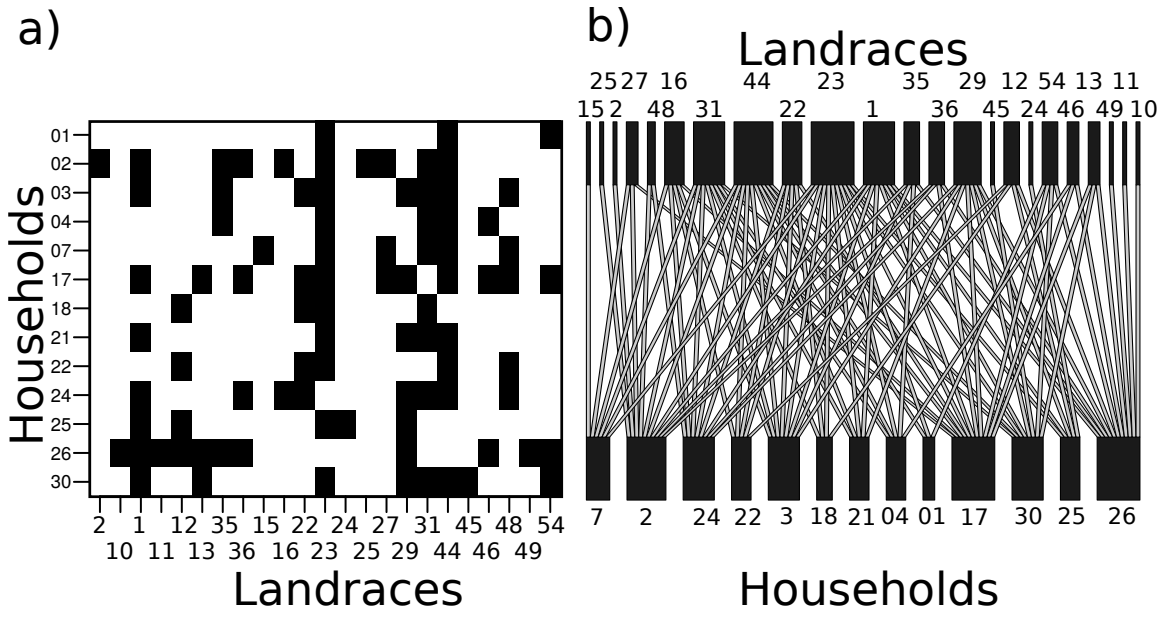

Figure 2: a) Example of incidence matrix where 0 are black cells and 1 are white cells; b) example of bipartite network between households and landraces (dataset AB-M02).

Summing over plant species, the number of species cultivated by household i, S_i, is

$$S_i = \sum_j X_{ij}. \tag{1}$$

Summing over farmers, the number of households that cultivate plant j, F_j, is

$$F_j = \sum_i X_{ij}. \tag{2}$$

Quantities N, S_i, F_j and X_{ij} are finally linked by the following relations:

$$N = \sum_i S_i = \sum_j F_j = \sum_{i,j} X_{ij}. \tag{3}$$

Following the network terminologly, S_i is also called the household's degree and F_j the plant's degree.

4.2 Variability of households' and plants' degrees

4.2.1 Description of the test on degree distributions

First, we evaluate whether all households in the same village grow a similar number of species or if there is high heterogeneity between farms' species richness. Formally, we test whether the degrees F_i follow binomial distributions by considering a statistics T that compares the observed variance of the plants degree with the one that would have been expected if the degrees S_i were following independent and identically distributed Binomial distribution.

$$T_{\text{row}} := \frac{\widehat{\text{Var}}(S)}{n\hat{p}(1-\hat{p})},$$

where $\hat{p} := \frac{N}{nm}$ is the density of the incidence matrix and $\widehat{\text{Var}}(S) = \frac{1}{n-1}\sum_{i=1}^{n}\left(S_i - m\hat{p}\right)^2$ is the empirical variance of (S_i), $i = 1,\ldots,n$. Large T_{row} values suggest that the household's species richness is highly heterogeneous whereas small T_{row} values suggest more equity. The statistical significance of T is assessed by a parametric bootstrap method working as follows. For $i = 1,\ldots,n_{sim}$, a new incidence matrix $\mathbf{X}^{(i)}$ is generated by sampling independent Bernoulli distributions with parameters $\hat{p}$ in each entry. For all these matrices, the link density $\hat{p}^{(i)}$, the empirical variance of the household's degrees $\widehat{\text{Var}}^{(i)}(S)$ and the variance ratio $T_{\text{row}}^{(i)}$ are computed. Finally, the left p-value and right p-values are respectively

$$\text{pval}_{L,\text{row}} := \frac{\#\left\{i:\ T_{\text{row}}^{(i)} < T_{\text{row}}\right\}}{n} \quad \text{and} \quad \text{pval}_{R,\text{row}} := \frac{\#\left\{i:\ T_{\text{row}}^{(i)} > T_{\text{row}}\right\}}{n}.$$

The plants' degree distribution are evaluated in a similar fashion.

$$T_{\text{col}} := \frac{\widehat{\text{Var}}(F)}{m\hat{p}(1-\hat{p})}; \qquad \widehat{\text{Var}}(F) = \frac{1}{m-1}\sum_{j=1}^{m}\left(F_j - n\hat{p}\right)^2.$$

The corresponding p-values are also evaluated by parametric bootstrap. In our analysis, the parameter n_{sim} is fixed to 10000.

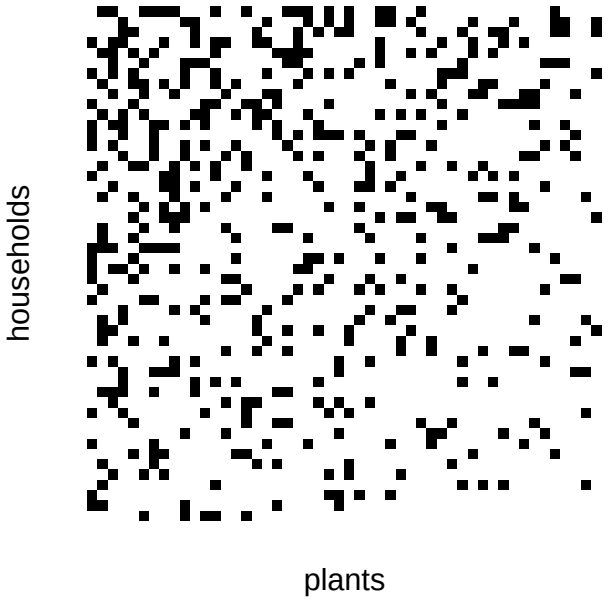

Figure 3: Incidence matrix with entries generated independently and identically distributed according to a Bernoulli distribution with probability 0.2.

Under a null model, called Erdős-Renyii, where all the entries of $\mathbf{X}$ follow independent Bernoulli distribution with identical parameter, the households' degrees and the plants' degrees follow binomial distributions. consequently, any small p-value ($\text{pval}_{L,\text{row}}$, $\text{pval}_{R,\text{row}}$, $\text{pval}_{L,\text{col}}$, $\text{pval}_{R,\text{col}}$) entail that this Erdős-Renyii model is not realistic.

4.2.2 Application of the test on degree distributions to a toy example

Figures 3, 4 and 5 display three examples of incidence matrices. The last two matrices were generated by assuming groups of plants and groups of households according to a Latent Block Model (see presentation in the next subsection). The households and the plants were sorted by degrees inside groups. Note that this structure of groups is generally unknown on real data set and has to be recovered by statistical inference techniques. In Figure 4, the incidence matrix was generated from i.i.d. Bernoulli random variables. Hence its row and column degrees follow binomial distributions. This corresponds to the null hypothesis of the test on the variance of degrees. The tests are non significant for this incidence matrix (see table 3). In Figure 4, some households were assumed to grow more plants than others and some plants assumed to be more popular. Therefore, as expected, the tests on the variance of degrees show clearly an over-dispersion for households

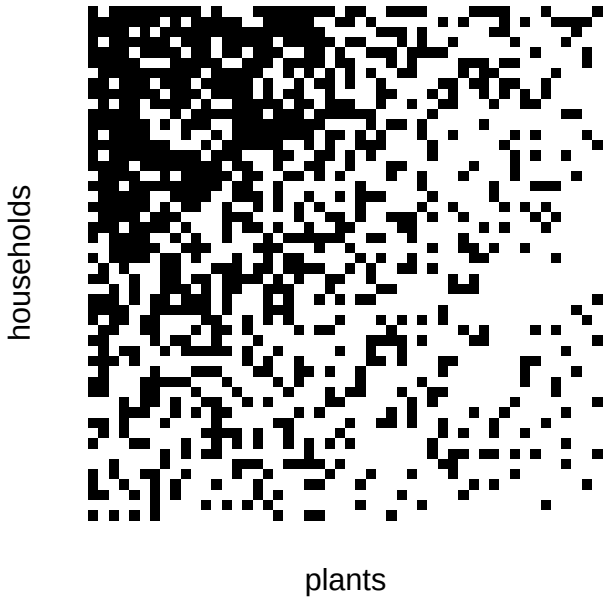

Figure 4: Incidence matrix generated with heterogeneous distribution for different groups of plants and households (see Figure 7 in next subsection for details). Some households grow more plants than other and some plants are more popular.

Figure 5: Incidence matrix generated with distribution implying particular association between plants and households (see Figure 6 in next subsection for details). Two groups of plants are mainly grown by corresponding subgroups of households.

and plants. In Figure 5, there exist particular associations between some groups of households and some groups of plants. However, the degree is quite homogeneous for households. For plants, a heterogeneity appears since the groups of households are not of the same size.

As illustrated on these three examples, the tests on the variance of degrees may detect heterogeneity but some particular structure of association may be missed as in the case of Figure 5. Indeed, the tests are performed independently on households an on plants and thus are not able to detect patterns of association.

4.3 Revealing data structure through latent block models

4.3.1 Description of the latent block models

In order to cluster the households and the plants simultaneously on the basis of the incidence matrix $\mathbf{X}$, we propose to use a probabilistic model called Latent Block Model (Govaert and Nadif, 2008; Keribin et al., 2014). It consists in assuming a mixture distribution both on the households and on the plants. According to this model, the network is generated according to latent blocks (also called clusters) of household and latent block of households. Conditioned to these latent blocks, the probability that a household i grows a plant j only depends on the block $V(i)$ to which household i belongs and the block W_j to which plant j belongs. For all $1 \leq i \leq n$, $1 \leq j \leq m$, $1 \leq q \leq Q$, $1 \leq l \leq L$, the probability that i belongs to block q, that j belongs to block l and the conditional probability of X_{ij} given the block V_i and W_j are respectively denoted

$$\begin{aligned} \mathbb{P}(V_i = q) &= \alpha_q, \\ \mathbb{P}(W_j = l) &= \beta_l, \\ \mathbb{P}(X_{ij} = 1 | V_i = q, W_j = l) &= \pi_{ql}, \end{aligned}$$

where $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_Q, \beta_1, \ldots, \beta_L, \pi_{11}, \ldots, \pi_{QL})$ is the vector of unknown parameters to be estimated under the obvious constraints $\sum_q \alpha_q = 1$, $\sum_l \alpha_l = 1$. This model is quite flexible since it can account

not only for situation where there is modularity i.e. to each block of households is associated a unique block of plants and these households tend to grow mainly plants from this block and very few from other blocks but also situations where there are richer households (growing significantly more plants than others) and/or more popular plants (grown by significantly more households).

The standard procedures to obtain maximum likelihood estimated when dealing with latent variables rely on the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). However, the computation of the conditional distribution of the latent variables with respect to the observed data is not tractable which makes the E step infeasible. Following Govaert and Nadif (2008), we use a variational approach to cope with this difficulty. The number of blocks of households Q and the number of blocks of plants L are chosen thanks to the integrated completed likelihood (ICL) criterion as proposed in Keribin et al. (2014). Once the parameters have been estimated, we obtain as a by-product the posterior probabilities $\mathbb{P}(V_i = q|\mathbf{X})$ and $\mathbb{P}(W_i = l|\mathbf{X})$ from which the true blocks are estimated. We can then provide a new representation of the incidence matrix $\mathbf{X}$ where the rows (households) and the columns (plants) have been reorganized in homogeneous blocks. We used the R package (Leger, 2015) to perform the estimations and the model selection.

4.3.2 Application of LBM to a toy example

Figures 6, 7 and 8 are illustrations of the block clustering provided by the LBM in three typical cases. The cases of Figure 6 and 7 are the same as those in Figures 5 and 4 respectively. The groups were considered as latent/unknown and the households and plants were clustered in homogeneous blocks by using the inference procedure described above. This is illustrated in Figure 6 where the same incidence matrix is plotted before and after re-organization according to the estimated blocks. In Figure 6, the difference between the two groups of households comes from the two last groups of plants. The first group of plants is equally grown up by households of any group. On the contrary, the second group of plants is mainly grown up by the second group of households and the third group of plants is mainly grown up by the first group of households. In Figure 7, the households can be separated on the basis on the number of plants that they grown up, a group can be said to be rich and the other to be poor. Similarly, two groups are also found for plants, one

Figure 6: Incidence matrix generated according to a LBM with 3 blocks of plants, 2 blocks of households and $\pi = \begin{pmatrix} 0.5 & 0.1 & 0.6 \\ 0.5 & 0.6 & 0.1 \end{pmatrix}$. Left: observed incidence matrix. Right: same incidence matrix re-organized and clustered in homogeneous blocks obtained by LBM inference.

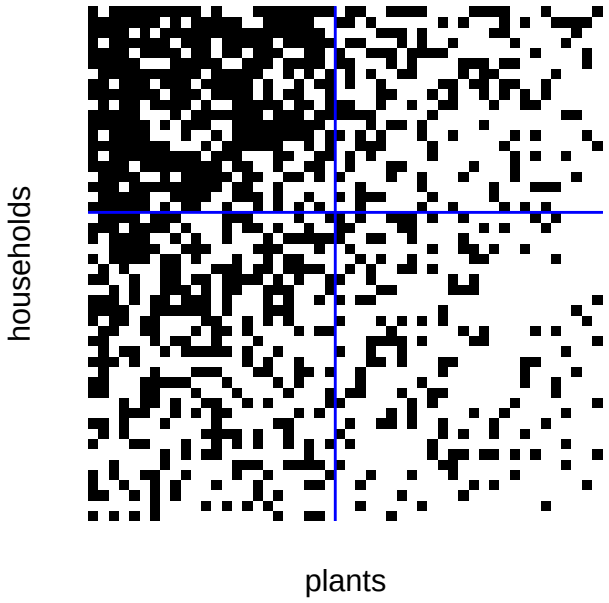

Figure 7: LBM clustering when the data are generated with 2 blocks of households (rich and poor households), 2 blocks of plants (rare and frequent plants) and $\pi = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.2 \end{pmatrix}$

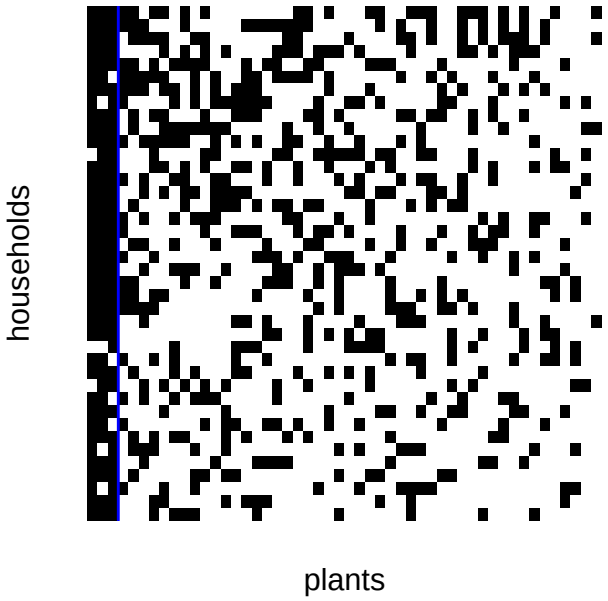

Figure 8: LBM clustering when the data are generated with 1 block of households, 2 blocks of plants (one block with only 3 plants) and $\pi = \begin{pmatrix} 0.9 & 0.3 \end{pmatrix}$

compounded of frequent / common plants and the other of rare plants. In Figure 8, households are similar and three plants are much more common than the others. Since the difference is quite clear and there are three plants, the ICL criterion for the LBM advocates for creating a block with only three plants. However, if there is only one or two outlier(s) or if the difference is less clear, this criterion may not separate this(these) outlier(s). This criterion for model selection is not designed for detecting outliers.

4.4 Uncovering outliers through principal component analysis

4.4.1 Configuration model

Fix the degree $(S_i)_{i=1,\dots,n}$ of each farm and $(F_j)_{j=1,\dots,m}$ of all plants in $\mathbf{X}$. The (bipartite) configuration model with parameters (S_i) and F_j is the uniform distribution over all incidence matrices that leave the degrees S_i and F_j unchanged. In the ecological literature, this model is sometimes referred as the Fixed-Fixed null model (Ulrich and Gotelli, 2012; Connor and Simberloff, 1979; Zaman and Simberloff, 2002). In contrast to the LBM, the configuration model takes for given that some households might grow much more plants than others and that some plants are more com-

mon than others, but apart from that the incidence matrix is sampled uniformly.

In order to simulate according to the configuration model, we use the tswap sequential algorithm (Miklós and Podani, 2004) implemented in the `permatswap` function of the *R* package `vegan`. The practitioner has to take a burnin and thinning parameters large enough so that the algorithm explores well the space of space of incidence matrices. Although the mixing time of tswap algorithm is unknown, the mixing properties of the sequence can be visually checked using the plot method of `permatswap`.

4.4.2 Principal Component analysis (PCA) on residuals

The expected incidence matrix under the configuration model with degrees (S_i) and (F_j) is denoted $\mathbb{E}_0[\mathbf{X}|(S_i,F_j)]$. Alternatively, $\mathbb{E}_0[\mathbf{X}|(S_i,F_j)]$ can be seen as the average over all permutations on the entries of $\mathbf{X}$ that keep the degree sequences for both plants and households unchanged. Then, the residual matrix $\mathbf{R}$ under the configuration model is the difference between the observed incidence matrix and its expectation under the configuration model

$$R_{ij} = X_{ij} - \mathbb{E}_0[X_{ij}|(S_i,F_j)] \tag{4}$$

If the incidence matrix $\mathbf{X}$ was drawn according to the configuration model, then $\mathbf{R}$ would have no particular structure. In order to check the absence of structure, we apply a (non-standardized) principal component analysis (PCA) on $\mathbf{R}$. As customary for PCA, the projection of the rows (*i.e.* the households) along the first principal directions allows (i) to uncover groups of households that effectively cultivate the same types of plants (ii) to detect outliers, that is households whose field plant composition is unusual when the effect of household richness has been removed. As an example, a household whose cultivated diversity is really high would not necessary be an outlier, but this household will be considered as an outlier if it does not grows some really common species. The projection of the columns of $\mathbf{R}$ along the first principal directions provides information on outlier species or groups of species.

4.4.3 Goodness-of-fit test of the configuration model

Before going further into the interpretation of the PCA, we need to test its statistical significance. This is also equivalent to testing whether $\mathbf{X}$ has been drawn according to the configuration model. Denote λ_{max} the largest singular value of $\mathbf{R}$ (*i.e.* the square-root of the largest eigenvalue of $\mathbf{R}^t\mathbf{R}$), we reject the null hypothesis when $\lambda_{\max}$ is unusually large compared to the one of $\mathbf{R}^P$ arising from permutations $\mathbf{X}^P$ of $\mathbf{X}$ leaving the degree of each row and each column invariant. Equivalently, this test rejects the null hypothesis when the largest eigenvalue in the scree plot is unusually large.

Under the null hypothesis, the matrix $\mathbf{R}$ is pure noise and all the singular values of $\mathbf{R}$ should be small. Under the presence of outliers or of a few groups of farms that preferentially cultivate some plants, the matrix $\mathbf{R}$ is expected to be the sum of a noisy component and a low-rank component measuring the deviance from the configuration model. As a consequence, the singular value of $\mathbf{R}$ should be higher under the alternative than under the null hypothesis.

Although calibrated differently, the largest singular value statistic has been fruitfully applied to other problems of community detection (Bickel and Sarkar, 2013).

4.4.4 A new representation of the incidence matrix

Ordering the households according to the coordinate of their projection along the first principal direction, we denote $\sigma_1(i)$ the farm index associated the i-th smallest coordinate. Similarly, $\sigma_2(j)$ stands for the reordering of the plants according to their projection on the first direction. These permutation (σ_1, σ_2) define a new representation $\mathbf{Y}$ of the incidence matrix:

$$Y_{ij} = X_{\sigma_1(i),\sigma_2(j)} \tag{5}$$

This provides an alternative visualization of the incidence matrix to the LBM.

4.4.5 Toy-examples

Let us describe three typical examples to understand the behavior of the above statistics. In all these examples, the number n of households is set to 40 and the number m of plants is set to 60.

First, we consider a model with degree heterogeneity. For each household $i = 1, \ldots, n$ and each

plants $j = 1, \ldots, m$, we draw independent uniform random variable a_i and b_j in $(0,1)$. Then, each entry X_{ij} is drawn according to a Bernoulli distribution with parameter $\min(2a_i b_j, 1)$. As a consequence, the incidence matrix $\mathbf{X}$ exhibits a large degree heterogeneity between households (resp. plants) with a low a_i (resp. b_i) value and households (resp. plants) with a high a_i (resp. b_i). It is therefore not unexpected that the LBM estimation procedure (Figure 9) recovers several groups of plants and household. The p-value of configuration model from Section 4.4.3 equals 0.39. Again, this is not surprising, since this incidence matrix has been sampled to a model similar to the configuration model. This implies that the block structure found by the LBM method only accounts for the degree heterogeneity. As the configuration model residuals are completely random here, both the PCA scree plot and the representation (eq. (5)) of the incidence matrix are uninformative. No household and no plants have outlier PCA coordinates (lower right panel).

In the second example, we draw the incidence matrix X as above. Then, we replace each entry of the first row by independent Bernoulli random variables with parameter 0.5. As a consequence, the first household is assumed to have a completely different behaviour from all the other household as it grows plants regardless of their scarcity (b_j) in the village. In Figure 10, the LBM representation is close to that of the first example. The p-value of the configuration test is smaller than 10^{-3}. This is corroborated with the fact that the scree plot exhibits an unusually large first eigenvalue. The first household is detected as an outlier by the first coordinate representation (lower-right panel). Finally, the PCA-based representation (upper-right panel) highlights the unusual behaviour of this household.

In the last example, we draw random variables a_i and b_j as above. Then, the households are divided in two groups A_1 and A_2 of size $n/2$ and the plants are divided in two groups B_1 and B_2 of size $m/2$. Then, the entry X_{ij} is drawn according to Bernoulli distribution with parameter $\min(p_{in} 2a_i b_j, 1)$ if $(i,j) \in A_1 \times B_1$ or $(i,j) \in A_2 \times B_2$ and parameter $\min(p_{out} 2a_i b_j, 1)$ if $(i,j) \in A_1 \times B_2$ or $(i,j) \in A_1 \times B_2$ with $p_{in} = 1.4$ and $p_{out} = 0.6$. Intuitively, the households from A_1 (resp. A_2) preferentially grow plants from B_1 (resp. B_2), but the model also allows the degree of the household and each plant to be heterogeneous inside the blocks. As a consequence, this model, called degree-corrected is neither a LBM with 2×2 blocks nor a configuration models but a blend

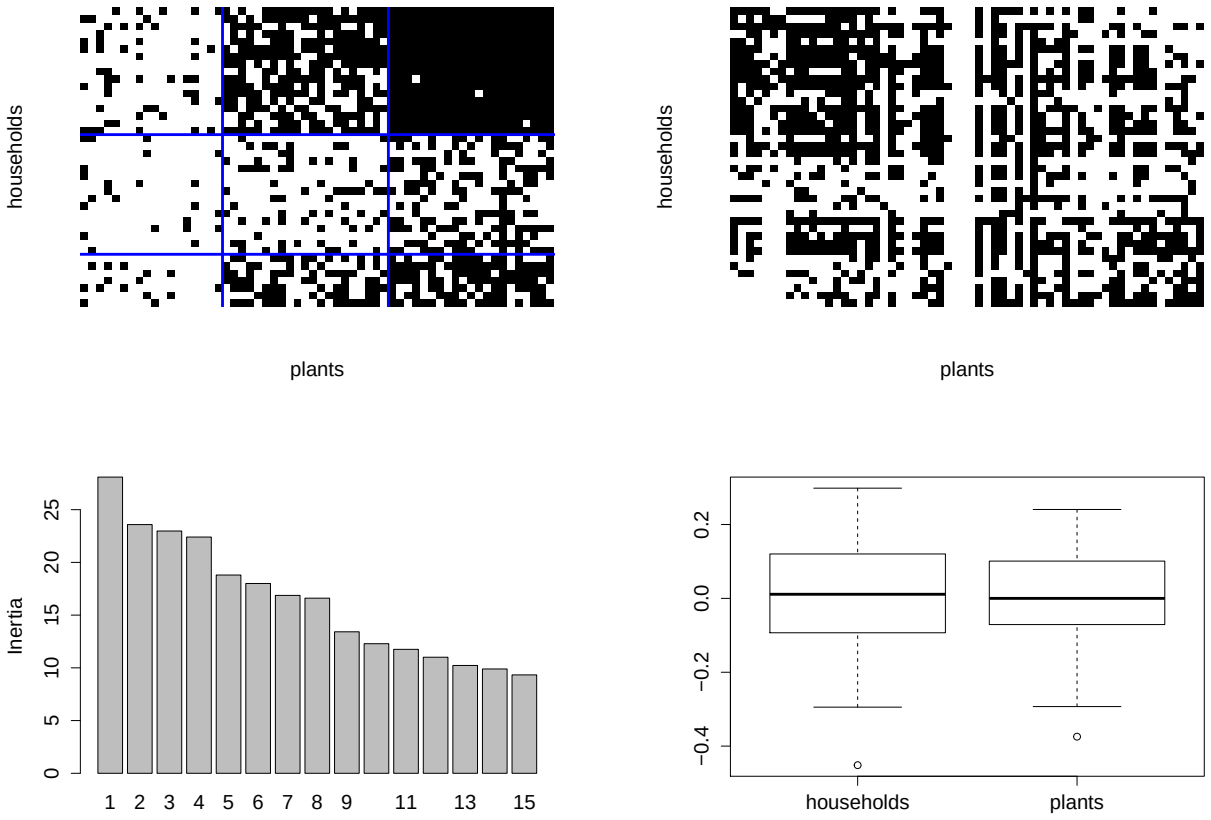

Figure 9: First example. The upper-left panel is the LBM representation. The lower left panel is the scree plot of the residuals PCA. The upper left panel is the representation of the incidence matrix according to the PCA ordering (5). The boxplots of the PCA first coordinates are pictured in the lower right panel.

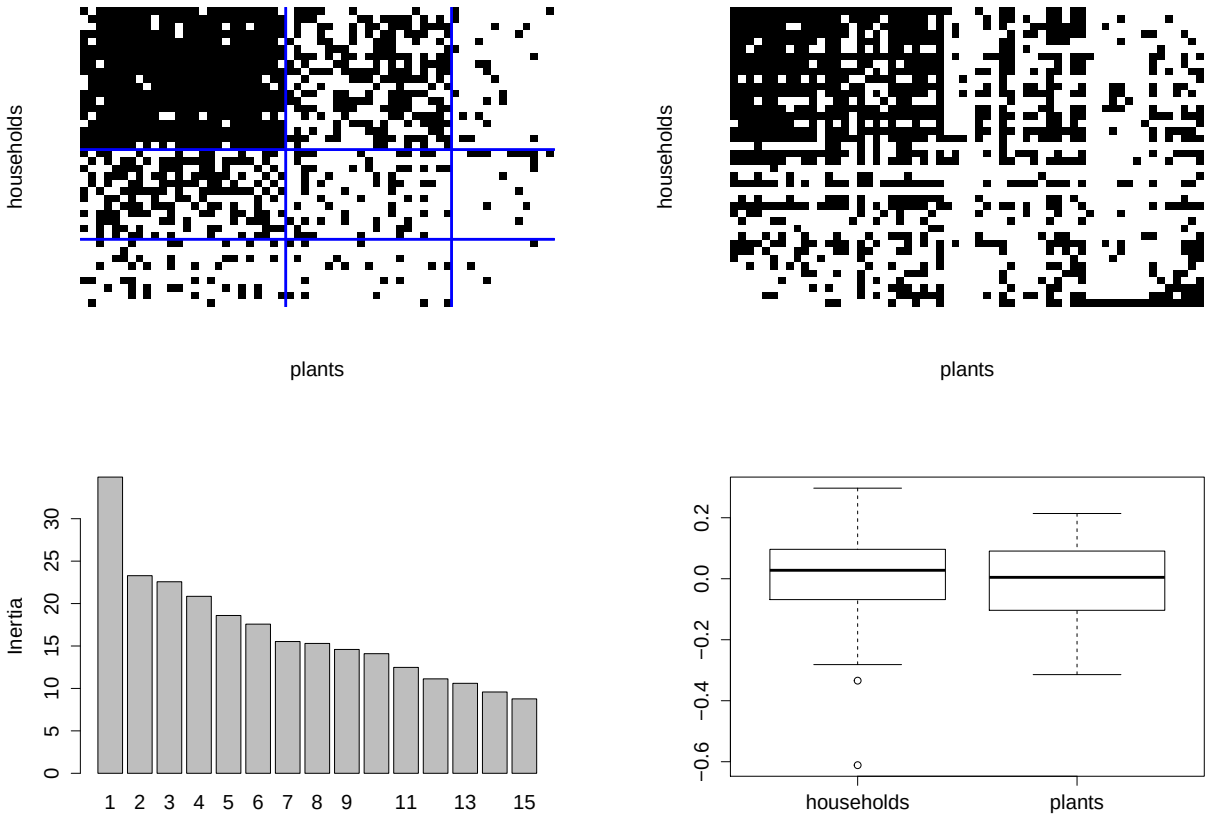

Figure 10: Second example. The upper-left panel is the LBM representation. The lower left panel is the scree plot of the residuals PCA. The upper left panel is the representation of the incidence matrix according to the PCA ordering (5). The boxplots of the PCA first coordinates are pictured in the lower right panel.

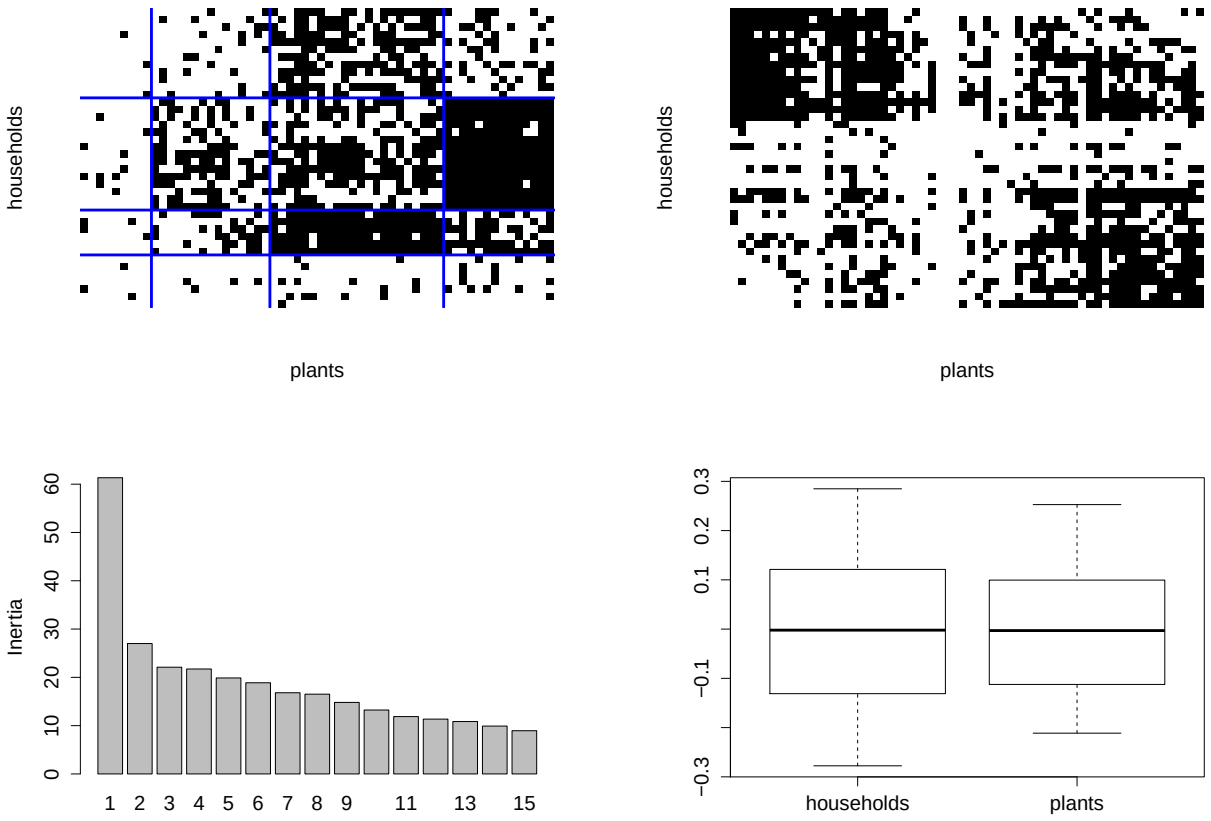

Figure 11: Third example. The upper-left panel is the LBM representation. The lower left panel is the scree plot of the residuals PCA. The upper left panel is the representation of the incidence matrix according to the PCA ordering (5). The boxplots of the PCA first coordinates are pictured in the lower right panel.

of them. The LBM estimation method recovers too many blocks (Figure 11) by grouping farms or plants being in the same group and having similar degrees. The p-value for the configuration test is found to be smaller than 10^{-3} (see also the scree plot). Contrary to the previous example, this unusually large singular values is not due to outliers (see lower-right panel) but to the presence of a block structure. The PCA-based matrix representation highlights the presence of these two groups of households and plants.

4.5 Measuring originality of households' contributions through diversity measures

We will now focus our attention on the distribution of cultivated plant diversity at the level of the sampled location (*e.g.* the village). As uncovered in the previous sections, some households might grow much more plant species than others (hence, the high variance in degree among households in the bipartite network). A question that remains unanswered is whether low-degree households contribute effectively more or less than high-degree households to the overall cultivated diversity - "effectively more" being understood as contributing more than expected if cultivated plants were chosen randomly from the pool of cultivated plants. In other words, the question is now whether low-degree households cultivate common plants only or contribute disproportionately to cultivated diversity by focusing only on plants that are cultivated by very few households.

4.5.1 Theoretical framework

Further expanding the notations introduced in subsection 4.1, we note p_{ij} the weight associated to the interaction between household i and plant j among all interactions of household i:

$$p_{ij} = \frac{X_{ij}}{S_i} \tag{6}$$

The proportion of all the connections in the network that are due to household i or plant j are respectively noted q_i and h_j:

$$q_i = \frac{S_i}{N} \tag{7}$$

$$h_j = \frac{F_j}{N} \tag{8}$$

We note H_i the diversity of plants cultivated by household i, as measured by Shannon entropy:

$$H_i = -\sum_j p_{ij} \log p_{ij} = \log S_i \tag{9}$$

The average diversity among households, weighted by their importance q_i, is noted H_α:

$$H_\alpha = \sum_i q_i H_i = \frac{1}{N} \sum_i S_i \log S_i \tag{10}$$

The diversity of plants cultivated by all households, when taken together and weighted by their importance q_i, is noted H_T and reads as:

$$H_T = -\sum_j \left[\sum_i q_i p_{ij} \right] \log \left[\sum_i q_i p_{ij} \right] = -\sum_j h_j \log h_j = \log N - \frac{1}{N} \sum_j F_j \log F_j \tag{11}$$

The difference between H_T and H_α is the turnover in diversity among households or β diversity, noted H_β:

$$H_\beta = H_T - H_\alpha = \log N - \frac{1}{N} \sum_j F_j \log F_j - \frac{1}{N} \sum_i S_i \log S_i \tag{12}$$

H_β can be further decomposed into individual turnover components, H_{iT}:

$$H_\beta = \sum_i q_i H_{iT} \tag{13}$$

where H_{iT} measures the "originality" of household i portfolio of plants when compared to the overall diversity of cultivated plants. An expression for H_{iT} can be found (Lande, 1996):

$$H_{iT} = -\sum_j p_{ij} \log \frac{S_i F_j}{N} \tag{14}$$

4.5.2 Measuring the diversity cultivated by plant-poor and plant-rich households

We now focus on measuring the evenness of cultivated by a subset I of households. More specifically, because we are interested in the subset of the most plant-poor or plant-rich households, we will assume that the set I contains all households belonging to a certain quantile of the distribution of S_i. The evenness of plants cultivated by households in set I is noted E_I and reads as

$$E_I = -\frac{\sum_j \left[\sum_{i \in I} q_{i,I} p_{ij} \right] \log \left[\sum_{i \in I} q_{i,I} p_{ij} \right]}{\log(m)}; \qquad q_{i,I} = \frac{S_i}{\sum_{i \in I} S_i} \,. \tag{15}$$

The evenness E_I is the diversity of plants cultivated by all households in set I divided by the logarithm of the total number m of type cultivated in the village. It measure the distribution's equity of species cultivated by households in I.

In order to assess whether the cultivated diversity is more even in plant-rich farms than plant-poor farms, we compare the value of $E_{Rich} - E_{Poor}$ to that of all realizations of the incidence matrix $\mathbf{X}$ under the configuration model (*i.e.* randomizing connections given degree sequences for both plants and households) by a permutation test.

4.5.3 Measuring the impact of plant-poor and plant-rich households

We now focus on measuring the β diversity $H_{\beta,I}$ due to the contribution of a subset I of households. As previously, the set subset I is made of the most plant-poor or plant-rich households. We can give an explicit formula for $H_{\beta,I}$ (Lande, 1996):

$$H_{\beta,I} = \sum_{i \in I} q_i H_{iT} = -\sum_i q_i \log q_i + \frac{1}{N} \sum_j \left[\sum_{i \in I} X_{ij} \right] \log\left(\frac{1}{F_j}\right) \tag{16}$$

The first term in the right-hand side of equation 16 relies on the expression of the α diversity $H_{\alpha,I}$ due to households in subset I:

$$H_{\alpha,I} = \frac{1}{N} \sum_{i \in I} S_i \log S_i = \frac{\sigma_I \log N}{N} + \sum_{i \in I} q_i \log q_i \tag{17}$$

where σ_I is the "volume" of interactions due to households belonging to subset I:

$$\sigma_I = \sum_{i \in I} S_i \tag{18}$$

The second term depend the correlation between a plant degree F_j and the number of households within the set I who possess this plant, noted $\varphi_{j,I}$:

$$\varphi_{j,I} = \sum_{i \in I} X_{ij} \tag{19}$$

Plugging equations 17, 18 and 19 into equation 16 yields the following expression for $H_{\beta,I}$:

$$H_{\beta,I} = \frac{\sigma_I \log N}{N} - H_{\alpha,I} - \frac{1}{N}\sum_j \varphi_{j,I} \log F_j \tag{20}$$

The quantity $D_I = \frac{1}{N}\sum_j \varphi_{j,I} \log F_j$ measures the deficit of originality displayed by the households in subset I that is due to their cultivation of "common plants".

Again, we assess the significance of $H_{\beta,I}$ by a permutation test based on the configuration model. As the set I contains all households belonging to a certain quantile of the distribution of S_i, all realizations of the incidence matrix $\mathbf{X}$ under the configuration model preserve the set of S_i values to be found in I. As a consequence, the quantity $\frac{\sigma_I \log N}{N} - H_{\alpha,I}$ in the right-hand side of equation 16 is invariant with respect to the configuration model. The quantity D_I in the right-hand side of equation 16, however, does not satisfy this invariance. Thus, large values of $H_{\beta,I}$ unusually large for the configuration model mean that households in subset I contribute more to cultivated biodiversity than expected by the number of types cultivated by households in I.

4.5.4 Measuring originality of households' contributions through diversity measures on toy examples

Model of simulation : Two groups of households are considered: rich (40% of households) and poor (60% of households). The plants are divided into two groups with same size: rare and popular. The entries of the incidence matrix are generated independently as Bernoulli random variables with probability p_{ij} (corresponding to household i and plant j) given by:

$$\text{logit}(p_{ij}) = \mu + \alpha(C_i) + \beta(K_j) + \gamma(C_i : K_j)$$

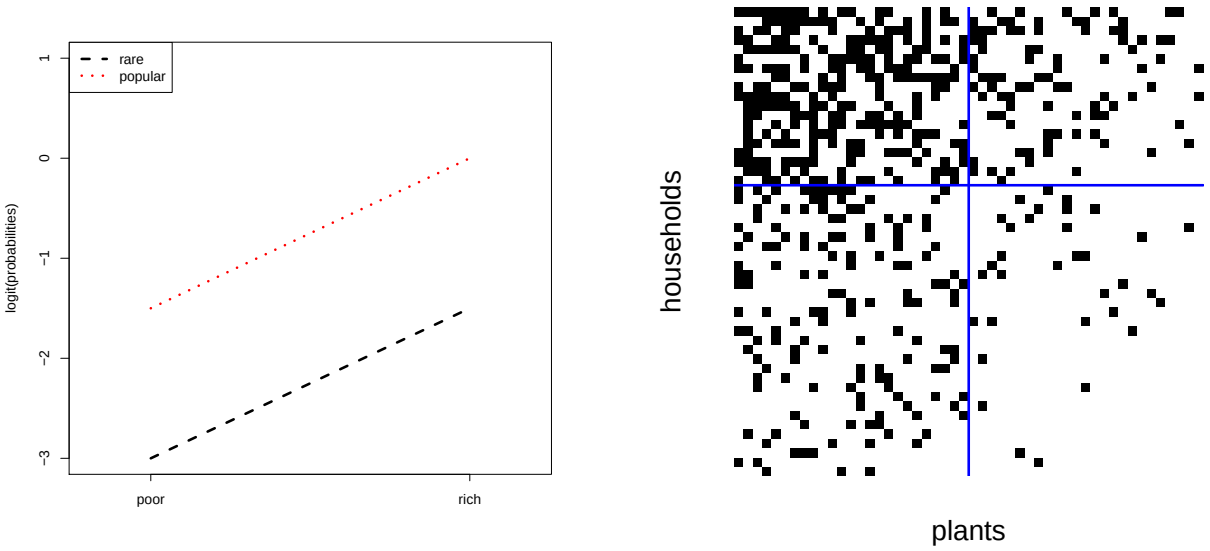

Figure 12: Toy example with equal contribution to diversity for rich and poor households. $\mu = -3$, $\alpha(rich) = \beta(popular) = 1.5$, $\gamma(rich, popular) = 0$. Left: probabilities for a household to grow a plant. Right: Incidence matrix.

where logit is the function $x \mapsto \log(x/(1-x))$, C_i indicates the group of household i, K_j the group of plant j and parameters μ, αs, βs, γs are chosen to lead to contrasted situations and such that

$$\alpha(poor) = \beta(rare) = \gamma(poor, rare) = \gamma(rich, rare) = \gamma(poor, popular) = 0$$

$$\alpha(rich) > 0, \beta(popular) > 0, \gamma(rich, popular) \neq 0$$

$$\alpha(rich) + 0.5 \cdot \gamma(rich, popular) > \alpha(poor) = 0$$

$$\beta(popular) + 0.40 \cdot \gamma(rich, popular) > \beta(rare) = 0$$

to ensure identifiability and coherence with regard to the modeled situations. The interaction term $\gamma(rich, popular)$ will then drive the respective contribution to diversity of rich and poor households. Indeed, if it is zero, the effect of being rich for growing a rare or a popular variety will be the same.

Three contrasted toy examples: Figures 12, 13 and 14 correspond respectively to the three following cases:

1. The rich and poor households have the same contribution to diversity with respect to their own richness. In the model of simulation, the interaction term $\gamma(rich, popular)$ was then

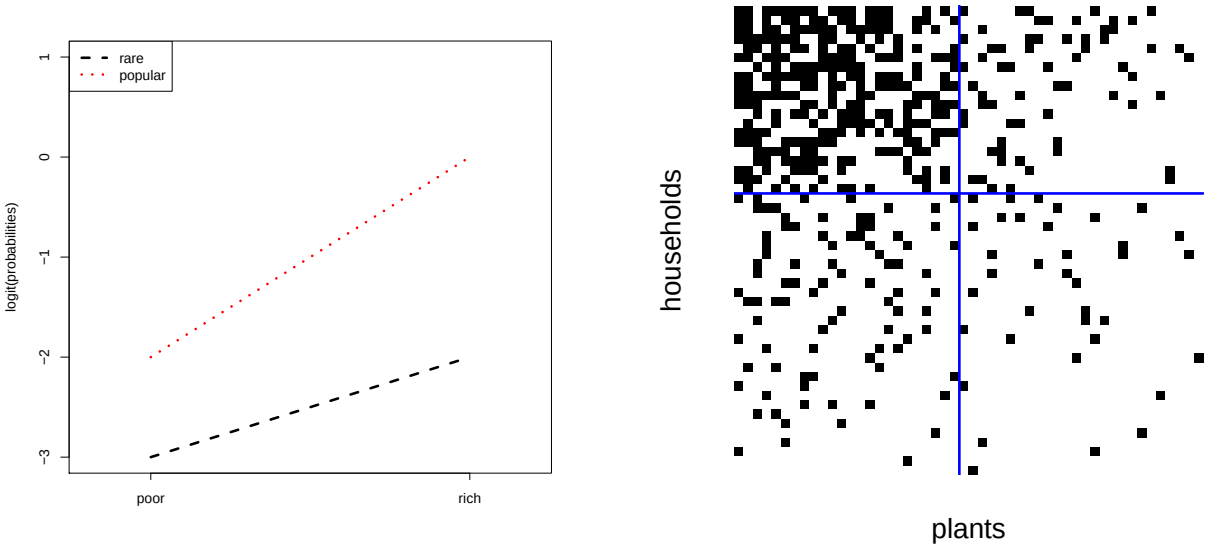

Figure 13: Toy example with greater contribution to diversity of poor households. $\mu = -3$, $\alpha(rich) = \beta(popular) = \gamma(rich, popular) = 1$. Left: probabilities for a household to grow a plant. Right: Incidence matrix.

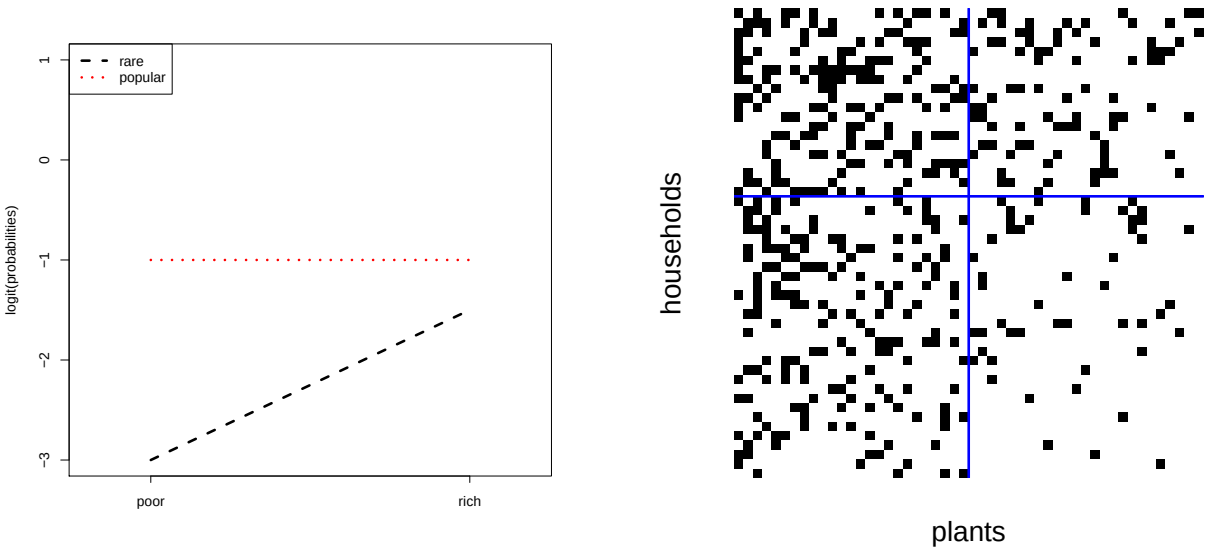

Figure 14: Toy example with greater contribution to diversity of rich households. $\mu = -3$, $\alpha(rich) = -\gamma(rich, popular) = 1.5$, $\beta(popular) = 2$. Left: probabilities for a household to grow a plant. Right: Incidence matrix.

fixed to 0.

2. The poor households have a greater contribution to diversity since they grow with nearly the same probability rare plants and popular plants while this probability of growing rare plants is clearly smaller than the probability of growing popular plants for rich households.

3. The rich households have a greater contribution to diversity. They are richer since they have the same ability of growing rare and popular plants.

Results in Table (P.: tableau sous odt ?) are coherent with what was expected when simulating. For the first case, nothing was found significant. For the two other cases, the tests on evenness and on the contribution to diversity of rich and poor households agreed. Indeed, for instance, in case of Figure 13, the rich households are found to contribute less than expected to diversity (null hypothesis rejected on left side), the poor households are found to contribute more than expected to diversity (null hypothesis rejected on right side) and the difference of evenness is found significantly smaller than expected (null hypothesis rejected on left side).

5 Patterns of local crop diversity: results of the meta-analysis

The tests performed in the meta-analysis are summarized in Table 5 and 6 for species and infra-species diversity, respectively.

5.1 Variability of households' and plants' degrees

Two null hypotheses (H_0) are tested in this section: 1)species and infra-species diversity is randomly distributed among households from the same village (homogeneity of the household degrees); 2) crop richness is randomly distributed within the same village (homogeneity of the plant degrees). More specifically, the aim of this section is to detect the existence of over-dispersion (significant test on the right) or under-dispersion (significant test on the left) of degree distribution for households and plants, respectively.

Species diversity For households, H_0 was rejected on the right side (16 times over the 18 tested datasets) for the variability of households' degree (Table 3). There was only one case where the test was not significant on both sides (SC-M05) and one case where the test was rejected on the left side (CL-M01). These results indicate that the number of species grown per household from the same village is generally over dispersed with few households growing more species than expected. For the variability in degree of species, this pattern was even stronger, with a systematically over-dispersed degree distribution.

Infra-species diversity For households at the infra-specific level, the pattern is completely different as H_0 is rejected on the right side only 3 times over the 32 tested datasets (ME-M01, SC-M04, SC-M07), and 11 times on the left side (Table 4). These results indicate an under-dispersion of the degree distribution when we consider the distribution of landraces at the village scale. For degree of landraces, H_0 is mostly rejected on the right side with 29 times over the 32 data sets, indicating, as for the species level, an over-dispersion of the degree distribution.

5.2 Structure detection through model-based clustering (LBM)

This section aims at detecting the existence of a structure within inventory datasets at the village scale using LBM, a model-based clustering approach.

Species diversity The clustering method applied on the different datasets detected from one to three clusters for the households and from two to three clusters for the species. These results are similar to the 'toy' example illustrated in Figure 7. Therefore, the clustering seems mostly driven by the heterogeneity in degree of both households and species. Households were clustered together because they grow almost the same species. In the case of two clusters for households, we then define the 'plant-poor' household cluster as the one with the lower density and the 'plant-rich' household cluster as the one with the higher density. In the case of two groups for the species, we define the 'rare species' cluster as the one with the lower number of links and the frequent species cluster as the one with the higher number of links.

Infra-species diversity The clustering method detected from one to two clusters for the households and from one to four clusters for landraces. For four datasets (DJ-M039a, DJ-M045b, DJ-M045c, DJ-M045d), only one cluster was detected both for households and landraces (Table 4). These results of low clustering are consistent with the low variability of the degrees both for the households and the landraces observed in section 5.1. Similarly, 26 additional data sets with an under-disperse had only one block for the households. These findings indicate that for landraces diversity, a lower heterogeneity is generally observed among households with almost the same landraces grown per household. Only three data sets showed two blocks for the households (ME-M01, SC-M04, SC-M07). Nevertheless it is still possible to distinguish between frequent and more rare landraces.

5.3 Outlier detection through PCA

We then used a Principal Component Analysis (PCA) to detect "outliers" in addition to the plant-poor and plant-rich households identified previously.

Species diversity Using the test introduced in section 4.4.3, H_0 was rejected 9 times over the 18 datasets at the $\alpha = 0.05$; rejecting H_0 highlights the existence of outliers. These outliers are generally two or three per dataset and can be characterized as households that grow a different subset of species compared to other households with an equivalent degree, *i.e.* belonging to the same cluster.

Infra-species diversity H_0 was rejected for only four datasets over the 32 datasets (CL-M02, DJ-M018a, DJ-M018b, DJ-M030). These results indicate that in addition to growing almost the same number of landraces, households from the same village grow globally the same portefolio of landraces. Note that for these four datasets, only one cluster was detected with the LBM (CL-M02, DJ-M018a, DJ-M018b, DJ-M030). Therefore, in this case we have households growing a particular subset of landraces and having an equivalent degree.

5.4 Households' contributions to local diversity

In the analyses reported on in this section, households were arbitrarily separated into 'plant-rich' households and 'plant-poor' households. Evenness (E) and contribution (H_β) were computed for each of these two groups.

Species diversity The tests on the difference between E_{rich} and E_{poor} revealed that plant-rich households had a significantly higher evenness in five cases (CL-M01, OC-M04, OC-M07, OC-M11, OC-M12). The group of plant-poor households contributed significantly more than that of plant-rich households in only one case (EG-M08). H_0 was not rejected in the other cases, indicating no significant difference in terms of contribution to the global diversity by the plant-rich group of households compared to the plant-poor group.

Our findings on the difference between E_{rich} and E_{poor} converge with the test of the contribution of plant-rich and plant-poor households. Indeed, in five cases when the first test was significant on the right side (*i.e.* a significantly higher contribution to the global diversity by the plant-rich households than the plant-poor households), we observed that some plant-rich group contributed significantly to the global diversity and that some plant-poor group contributed significantly less than expected in four times of the five cases (Table 3). Two additional datasets showed a significant contribution of the plant-rich households (OC-M14 and SC-M05) and one additional dataset showed that the plant-poor households contributed significantly less than expected (OC-M13). The plant-poor households contributed significantly more than expected in only two cases. In one of these cases (EG-M05), the result is consistent with that of the test on evenness. In the other case (EG-M08), plant-poor households only showed a significant contribution to global diversity and not to evenness (EG-M08).

Infra-species diversity The tests of the difference between E_{rich} and E_{poor} households revealed that plant-rich households had a significantly higher evenness in six cases (DJ-M003a, DJ-M012a, DJ-M012b, DJ-M018a, DJ-M030, DJ-M036). H_0 was not rejected in the other cases, indicating no significant difference in evenness between plant-rich and plant-poor households. These results were not always convergent with the results of the tests dealing with the contribution of the plant-

rich and plant-poor household groups to diversity at the village level. Indeed, these latter tests gave convergent results (a significant contribution of few plant-rich households to the global diversity) in only two cases (DJ-M018a and DJ-M036) of the six in which the evenness difference was significant. In one additional dataset, few households from the plant-rich group contributed significantly less than expected (JW-M07). In one additional dataset, the plant-poor households contributed significantly more than expected (DJ-M012a). In three additional datasets, few households from the plant-poor group contributed significantly less than expected (CV-M02, DJ-M030, JW-M08). Finally, in two datasets, the plant-poor households contributed significantly more than expected (DJ-M0015a, SC-M07).

6 Discussion

6.1 Contrasted patterns of local crop diversity at the species and infra-species levels

Applying a set of network-based methods on a meta-data set of crop diversity reveals distinct sources of heterogeneity in terms of crop distribution at the local scale: 1. crop diversity among households is generally more heterogeneous at the specific level than at the infra-specific level; 2. heterogeneity in households' degrees is one explanation for this heterogeneity with blocks of low diversity households and of high diversity households (the same pattern is observed for species and landraces with blocks of common plants and blocks of rarer plants); 3. outliers households with original portfolios are another source of heterogeneity; 4. finally, depending on the circumstances, either low diversity or high diversity households can contribute disproportionately to local diversity by growing rare varieties.

These general results suggest two main explanations: heterogeneity in data collection methods and a diversity of socio-ecological and environmental contexts. As datasets were collected following different protocols, differences in sampling effort could have an influence on the observed diversity (Perrault-Archambault and Coomes, 2008). Nevertheless, a subset of the data sets for landraces were collected in the context of a coordinated global partnership of researchers

636 in order to use the same protocol and the same sampling strategy during data collection (Jarvis
637 et al., 2008), and datasets collected in this context also show different patterns (DJ-M012a, DJ-
638 M012b, DJ-M015a, DJ-M015b, DJ-M018a, DJ-M018b, DJ-M030, DJ-M036, DJ-M039a, DJ-M045b,
639 DJ-M045c, DJ-M045d). Consequently, variation in the agro-ecological and the socio-cultural con-
640 texts, and in interactions between these contexts, is likely to strongly shape the distribution of local
641 crop diversity.

642 More precisely, we observe that the findings of over-dispersion of the degrees at the specific
643 level and an under-dispersion at the infra-specific level is strengthened by the results of classifi-
644 cation using LBM. Indeed, in the cases of over-dispersion, two or three blocks of households are
645 detected whereas for cases of under-dispersion, only one block of households is detected. Con-
646 vergence of the results between these two approaches indicates that the variability of the degree
647 distribution is probably the main driver of block structure. It thus makes sense to use as null
648 model a configuration model, controlling for degree, for the following tests because this would
649 allow assessment of whether other structural drivers in addition to the degree overcome to shape
650 the patterns of diversity. From an ethnobiological or agroecological point of view, the block de-
651 tection means that households can be distinguished according to the level of diversity they grow.
652 We identify high diversity and low diversity households. Similarly, for plants, we identify com-
653 mon species/landraces (present in fields of most households) and rare species/landraces (grown
654 by few households). Such patterns in terms of distribution of local crop diversity are quite com-
655 mon in the literature and consistent with the findings of Jarvis et al. (2008), who find that growing
656 area and landrace diversity are related.

657 From an ethnobiological point of view, these findings reflect the fact that ways of managing
658 diversity differ between the specific (crop species) and the infra-specific levels (landraces). Grow-
659 ing numerous species is more complicated than growing numerous landraces, for several reasons.
660 First, each species has its specific needs in terms of soil quality and preparation, sowing date,
661 quantity of labour required and when it is required, and so on (Garine and Raimond, 2005). Among
662 landraces of the same species, these needs are not so divergent. Households possessing a relatively
663 large land holding area have more chance to encounter different soil types and quality among
664 their fields. Also, larger households or those with an extensive social network can expect to have

665 an adequate labour supply (Abizaid et al., 2015) to grow a large portfolio of species (Garine and
666 Raimond, 2005). Thus, farmers with more assets and labor tend to cultivate a larger field area and
667 have greater crop diversity (Zimmerer, 1991; Coomes and Ban, 2004; Alvarez et al., 2005). Small-
668 holder poverty may limit the diversity of crops that can be raised. Previous studies concluded that
669 certain species are needed to meet basic needs (*e.g.,* food, medicinal, etc.) and other species are
670 more optional, reflected by higher levels of infra-specific diversity for food stables compared to
671 other crops (Jarvis et al., 2008), especially under stressful abiotic conditions (Labeyrie et al., 2013).
672 Another possible explanation of the lower heterogeneity for degrees for landraces is that several
673 landraces of the main species may be grown to fill diverse needs driven by cultural and dietary
674 preferences, shifts in market demand and labour availability (Brush and Meng, 1998; Gauchan
675 et al., 2005; Johns et al., 2013), heterogeneity in soil and water resources (Bisht et al., 2007; Bellon
676 and Taylor, 1993), biotic stresses (Finckh and Wolfe, 2006), and the need to enhance pollination
677 levels via outcrossing (Kremen et al., 2002). Much infra-specific diversity is held at the community
678 rather than within individual households (Mulumba et al., 2012; Brush et al., 2015; Fenzi et al.,
679 ress). In addition, in agroecosystems where many species are grown, households maintaining col-
680 lections of landraces will be few because less varietal diversity of the crop species is available to
681 the farmer due to financial, social or policy constraints. Finally, the greater heterogeneity of crop
682 diversity at the specific level compared to the infra-specific level may lie in the traits of the crop
683 species considered in the analysis and their reproductive systems In their broad comparison of
684 nomenclature systems Jarvis et al. (2008) showed that farmers use more detailed classifications for
685 clonally reproduced crops than for inbreeders, partial outbreeders or outbreeders. This hypoth-
686 esis was confirmed in our dataset. The only cases where over-distribution of household degree
687 was observed at the infra-specific level (ME-M01, SC-M04, SC-M07) were all villages in which the
688 staple food was provided by clonally propagated species (manioc, taro).

689 We applied additional tests to detect more detailed patterns in crop diversity within the meta-
690 data set and the sources of divergence in terms of crop portefolio composition. Our analysis of
691 outliers identified certain households holding unique portefolios of species or landraces. In most
692 cases, it is the high diversity households that mainly contribute to the global diversity. These find-
693 ings are consistent with the hypothesis of nestednes and of a sink-source dynamics described in

Alvarez et al. (2005) and Coomes (2010), and frequently postulated importance, in the dynamics of local diversity, of one or a small number of experts or nodal farmers in a village (Perrault-Archambault and Coomes, 2008; Boster, 1983; Padoch and Jong, 1991; Peroni and Hanazaki, 2002; Salick et al., 1997; Subedi et al., 2003; Tapia, 2000).

Nevertheless, it would be incorrect to say that it is a consistent tendency in the meta-data set. Indeed, we observed the opposite relationship in other data sets whereby low diversity households contributed significantly to the local diversity (EG-M05, EG-M08, DJ-M015a, SC-M07). In some case, one or a few farmers grow rarer species or landraces owing to curiosity, for aesthetic reasons, or to maintain a social status of expert at the local level (Elias et al., 2000; Meilleur, 1998; Hawkes, 1983), or to have an object that the others do not have (Coomes and Ban, 2004). Possessing an infrequent species or landrace might, for instance, allow a young farmer to distinguish himself from others in societies independently of economic capital. Having an object that others do not have, could increase the value of the eventual transfer to other members of the community. Cultivating rarity helps both to gain a social status within the village and to have highly valued objects to exchange (Caillon and Lanouguère-Bruneau, 2005). Additional factors influence the distribution of local crop diversity, for instance, the role played by differences associated with gender and generation, access to local seed markets, farmers' food preferences, and the market value of crops. Vertical transmission of seeds and knowledge occurs between mothers-in-law and daughters-in-law in patrilinear societies with virilocal residence rules, and constituting another source of divergence in crop diversity between families in the same neighbourhood (Labeyrie et al., 2013).

More generally, because these distinct patterns of crop diversity have been detected in different agro-ecological environments and socio-cultural contexts without controlling for other potential controlling factors (and without additional information about each village), it is not possible to assess how one particular agro-ecological environment and socio-cultural context shapes the distribution of local crop diversity. Additional studies are needed in this direction to detect the local drivers influencing the observed distribution of crop diversity.

6.2 Relevance of the network-based methods

The network-based methods introduced in this paper provide a set of useful tools to analyses the distribution of local diversity in crop species and varieties. Indeed, our framework allowed us to answer four key questions:

1. whether households' and plants' degrees are more variable than expected under a null model proposing a homogeneous probability of interaction between potential partners;
2. whether household-plant interactions are structured by blocks and, if so, what these blocks are;
3. whether certain plants or certain households behave as obvious outliers in their pattern of interactions;
4. whether low-degree and high-degree households contribute significantly more or less than expected under a null model (the configuration model) to the overall diversity of plants cultivated locally.

The combination of these different indices, tests and metrics provides a realistic and complete picture of the complex structure of crop diversity. For instance, this framework readily detected cases in which plant diversity is different in two different villages (through the latent-block models) and identified households – be they low-degree or high-degree households – as unique and important providers of plant diversity (through PCA outlier uncovering and measures of uniqueness).

One strength of this framework is the use of a hierarchy of null models with increasing complexity. For instance, the most simple model for a bipartite network with variable degrees is the Erdős-Rényi $G(N, p)$ model restricted to interaction between nodes from the two different categories (each link has the same probability of occurring). Deviations from this null model allow assessment of degree heterogeneity or the presence of blocks (groups of households that preferentially cultivate a certain group of species). When looking for more elaborate structures in the network (and not only degree distributions), we relied on the configuration model, which randomizes interactions while keeping all degrees in the network fixed. Consequently, one can dis-

entangle whether the observed patterns, such as the block structure, are simply explained by the degree heterogeneity or are truly emergent properties.

It is important to note that our approach can be extended to other datasets from other disciplines, including ecology, to detect particular patterns in bipartite networks. In ecology, the tests could efficiently supplement metrics that are routinely used, such as modularity or nestedness scores (Fortuna et al., 2010). Depending on the size of the dataset, latent-block models can be as informative (or more) as traditional modularity-computing techniques in finding underlying structures within bipartite datasets (Leger, 2015). Moreover, LBMs can also elucidate non-modular blocks such as quasi-partite structures (*i.e.*, when such structures are not exactly bi- or multi-partite but quite close) within a network. Of course, the power of all such methods depends heavily on the number of nodes in the network, but the application to ecological questions of the set of methods proposed here could readily generate much more informative descriptions of ecological networks than connectance, modularity and nestedness scores alone.

The approach used in this paper does not rely on a direct estimation of nestedness because the different methods available to compute nestedness do not converge (Supplementary material Fig 15). However, the set of methods designed here to uncover the uniqueness of contributions to diversity of plant-rich and plant-poor households actually provide complementary information on whether specialists interact preferentially with generalists or not, as assumed under a "nested" scenario in ecology. We thus suggest that this toolkit could be used as an alternative to the classic methods of nestedness detection usually applied to ecological datasets (Podani and Schmera, 2012).

From a methodological point of view, the configuration model must be accompanied by several caveats. Most prominently, the fact that the degrees of all nodes are fixed makes the model highly constrained. Relaxing the requirement that all samples of the models reproduce exactly the desired degrees, Chung and Lu (2002a,b) developed a model that generated graphs with given expected degrees; degrees of networks sampled from this model are allowed to vary slightly around a fixed expected value. Interestingly, the Chung-Lu model has recently been extended into the so-called degree-corrected stochastic block model (Karrer and Newman, 2011) incorporating both degree-heterogeneity parameters as in the Chung-Lu model and a block structure as in the LBM.

Such models would allow disentangling the households' overall crop richness and plant rarity from the preferences of certain households for specific groups of species (block structure). Inference methods for this model have been recently developed (*e.g.* Lei et al., 2014). However, the complexity of these models makes the estimation (and the computation of p-values) unreliable for small networks such as those considered in this study. Nevertheless, the Chung-Lu model and degree-correcting stochastic block models are promising directions of research and analysis of larger-scale ecological networks.

7 Conclusion

In this paper we develop new network-based indicators and statistical tests to characterize patterns of crop diversity at local scales. We applied this methodological framework to a meta-data set from 10 countries containing inventory data at the specific or infra-specific level. Our results identify different sources of heterogeneity local crop diversity: 1. diversity at the specific level is generally much more heterogeneous among households compared to diversity at the infra-specific level; 2. two or more groups of households can be identified based on their unique crop richness; 3. although diversity rich households often contribute most to global diversity, in some cases diversity poor households contribute rare species and varieties. . This analysis reveals the absence of any general pattern of crop diversity independent of agro-ecological and socio-cultural contexts, suggesting the need for further empirical research. Our methodological framework provides a useful approach and an informative overview of patterns in the distribution of diversity. The toolkit developed and applied in this study offers an alternative approach to the classic methods of nestedness detection in both ethnographic and ecological datasets.

Acknowledgements

The authors gratefully acknowledge the French Fondation pour la Recherche sur la Biodiversité (FRB) that made possible NetSeed, an international collaboration of researchers studying farmer seed networks. Mathieu Thomas was provided with a six month post-doctoral fellowship within this program. The Centre de Synthèse et d'Analyse sur la Biodiversité (CESAB) provided essential logistical support for regular workshops on the subject enabling us to develop and mature both

the theoretical and methodological aspects of our research. Additional support of the ongoing research collaboration through the MIRES and MADRES networks was provided by the following agencies: the Réseau National des Systèmes Complexes (RNSC), the Institut National de la Recherche Agronomique (INRA) and the Centre National de la Recherche Scientifique (CNRS). We are most thankful for the NETSEED researchers not directly include in the preparation of this paper for the fruitful discussions in which they participated during the project's different meetings.

References

ABIZAID, C., COOMES, O. T., TAKASAKI, Y., AND BRISSON, S. 2015. Social Network Analysis of Peasant Agriculture: Cooperative Labor as Gendered Relational Networks. The Professional Geographer 67:447–463.

ALMEIDA-NETO, M., GUIMARÃES, P., GUIMARÃES, P. R., LOYOLA, R. D., AND ULRICH, W. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. Oikos 117:1227–1239.

ALVAREZ, N., GARINE, E., KHASAH, C., DOUNIAS, E., HOSSAERT-MCKEY, M., AND MCKEY, D. 2005. Farmers' practices, metapopulation dynamics, and conservation of agricultural biodiversity on-farm: a case study of sorghum among the Duupa in sub-sahelian Cameroon. Biological Conservation 121:533–543.

BELLON, M. R. AND TAYLOR, J. E. 1993. "Folk" soil taxonomy and the partial adoption of new seed varieties. Economic Development and Cultural Change 41:763–786.

BIANCHI, F., BOOIJ, C. J. H., AND TSCHARNTKE, T. 2006. Sustainable pest regulation in agricultural landscapes: a review on landscape composition, biodiversity and natural pest control. Proceedings of the Royal Society B: Biological Sciences 273:1715–1727.

BICKEL, P. J. AND SARKAR, P. 2013. Hypothesis testing for automated community detection in networks. arXiv preprint arXiv:1311.2694 .

BISHT, I., MEHTA, P., AND BHANDARI, D. 2007. Traditional crop diversity and its conservation on-farm for sustainable agricultural production in Kumaon Himalaya of Uttaranchal State: A case study. Genetic Resources and Crop Evolution 54:345–357.

BOSTER, J. 1983. A comparison of the diversity of Jivaroan Gardens with that of the tropical forest. Human Ecology 11:47–68.

BOSTER, J. S. 1985. Selection for perceptual distinctiveness: Evidence from Aguaruna cultivars of *Manihot esculenta*. Economic Botany 39.

BRUSH, S. B., BELLON, M. R., HIJMANS, R. J., RAMIREZ, Q. O., PERALES, H. R., AND ETTEN, J. V. 2015. Assessing maize genetic erosion. Proceedings of the National Academy of Sciences 112:E1–E1.

BRUSH, S. B. AND MENG, E. 1998. Farmers' valuation and conservation of crop genetic resources. Genetic Resources and Crop Evolution 45:139–150.

838 CAILLON, S. AND LANOUGUÈRE-BRUNEAU, V. 2005. Gestion de l'agrobiodiversité dans un village
839 de Vanua Lava (Vanuatu) : stratégies de sélection et enjeux sociaux. Le Journal de la Société des
840 Océanistes pp. 129–148.

841 CHUNG, F. AND LU, L. 2002a. The average distances in random graphs with given expected de-
842 grees. Proceedings of the National Academy of Sciences 99:15879–15882.

843 CHUNG, F. AND LU, L. 2002b. Connected components in random graphs with given expected
844 degree sequences. Annals of combinatorics 6:125–145.

845 CONNOR, E. F. AND SIMBERLOFF, D. 1979. The assembly of species communities: chance or com-
846 petition? Ecology pp. 1132–1140.

847 COOMES, O. T. 2010. Of Stakes, Stems, and Cuttings: The Importance of Local Seed Systems in
848 Traditional Amazonian Societies. The Professional Geographer 62:323–334.

849 COOMES, O. T. AND BAN, N. 2004. Cultivated plant species diversity in home gardens of an Ama-
850 zonian peasant village in Northeastern Peru. Economic Botany 58:420–434.

851 CROWDER, D. W., NORTHFIELD, T. D., STRAND, M. R., AND SNYDER, W. E. 2010. Organic agricul-
852 ture promotes evenness and natural pest control. Nature 466:109–112.

853 DEMPSTER, A. P., LAIRD, N. M., AND RUBIN, D. B. 1977. Maximum likelihood from incomplete
854 data via the em algorithm. Journal of the Royal Statistical Society, series B 39:1–38.

855 DIAMOND, J. 2002. Evolution, consequences and future of plant and animal domestication. Nature
856 418:700–707.

857 DÊLETRE, M., MCKEY, D. B., AND HODKINSON, T. 2011. Marriage exchanges, seed exchanges,
858 and the dynamics of manioc diversity. Proceedings of the National Academy of Sciences of the
859 United States of America 108:18249–18254.

860 ELIAS, M., RIVAL, L., AND MCKEY, D. 2000. Perception and management of cassava (*Manihot*
861 *esculenta Crantz*) diversity among Makushi Amerindians of Guyana (South America). Journal of
862 Ethnobiology 20:239–265.

863 FENZI, M., JARVIS, D. I., ARIAS-REYES, L. M., L. M., LATOURNERIE-MORENO, L., AND TUXILL, J. in
864 press. Longitudinal analysis of maize diversity in Yucatan, Mexico: influence of agro-ecological
865 factors on landraces conservation and modern variety introduction. Plant Genetic Resources
866 Characterization and Utilization .

867 FINCKH, M. R. AND WOLFE, M. S. 2006. Diversification strategies, pp. 269–307. In B. M. Cooke, D.
868 Jones, Gareth, and B. Kaye (eds.), The Epidemiology of Plant Diseases. Springer Netherlands.

869 FORTUNA, M. A., STOUFFER, D. B., OLESEN, J. M., JORDANO, P., MOUILLOT, D., KRASNOV, B. R.,
870 POULIN, R., AND BASCOMPTE, J. 2010. Nestedness versus modularity in ecological networks:
871 two sides of the same coin? Journal of Animal Ecology 79:811–817.

872 FRASER, J., ALVES-PEREIRA, A., JUNQUEIRA, A., PERONI, N., AND CLEMENT, C. 2012. Convergent
873 adaptations: bitter manioc cultivation systems in fertile anthropogenic dark earths and flood-
874 plain soils in Central Amazonia. PloS one 7:e43636.

GARINE, E. AND RAIMOND, C. 2005. La culture intensive fait-elle disparaître la biodiversité ? In Dynamique de la biodiversité et modalité d'accès aux milieux et aux ressources, pp. 25–28, Paris, France. Institut Français de la Biodiversité.

GAUCHAN, D., SMALE, M., AND CHAUDHARY, P. 2005. Market-based incentives for conserving diversity on farms: the case of rice landraces in Central Tarai, Nepal. Genetic Resources and Crop Evolution 52:293–303.

GOVAERT, G. AND NADIF, M. 2008. Block clustering with bernoulli mixture models: Comparison of different approaches. Computational Statistics & Data Analysis 52:3233–3245.

HAWKES, J. G. 1983. The diversity of crop plants. p. 184pp.

HECKLER, S. AND ZENT, S. 2008. Piaroa manioc varietals: Hyperdiversity or social currency. Human Ecology 36:679–697.

JARVIS, D. I., BROWN, A. H., CUONG, P. H., COLLADO-PANDURO, L., LATOURNERIE-MORENO, L., GYAWALI, S., TANTO, T., SAWADOGO, M., MAR, I., AND SADIKI, M. 2008. A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. Proceedings of the National Academy of Sciences 105:5326–5331.

JOHNS, T., POWELL, B., MAUNDU, P., AND EYZAGUIRRE, P. B. 2013. Agricultural biodiversity as a link between traditional food systems and contemporary development, social integrity and ecological health. Journal of the Science of Food and Agriculture 93:3433–3442.

JORDANO, P. 1987. Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. American Naturalist 129:657–677.

JORDANO, P., BASCOMPTE, J., AND OLESEN, J. M. 2003. Invariant properties in coevolutionary networks of plant-animal interactions. Ecology Letters 6:69–81.

KARRER, B. AND NEWMAN, M. E. 2011. Stochastic blockmodels and community structure in networks. Physical Review E 83:016107.

KERIBIN, C., BRAULT, V., CELEUX, G., AND GOVAERT, G. 2014. Estimation and selection for the latent block model on categorical data. Statistics and Computing pp. 1–16.

KOLACZYK, E. D. 2009. Statistical analysis of network data: methods and models. Springer Science & Business Media.

KREMEN, C., WILLIAMS, N. M., AND THORP, R. W. 2002. Crop pollination from native bees at risk from agricultural intensification. Proceedings of the National Academy of Sciences 99:16812–16816.

LABEYRIE, V., RONO, B., AND LECLERC, C. 2013. How social organization shapes crop diversity: an ecological anthropology approach among Tharaka farmers of Mount Kenya. Agriculture and Human Values 31:97–107.

LANDE, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos pp. 5–13.

LAZEGA, E., MOUNIER, L., SNIJDERS, T., AND TUBARO, P. 2012. Norms, status and the dynamics of advice networks: A case study. Social Networks 34:323–332.

LECLERC, C. AND COPPENS D'EECKENBRUGGE, G. 2012. Social organization of crop genetic diversity. the g x e x s interaction model. Diversity 4:1–32.

LEGER, J.-B. 2015. blockmodels: Latent and Stochastic Block Model Estimation by a 'V-EM' Algorithm. R package version 1.1.1.

LEI, J., RINALDO, A., ET AL. 2014. Consistency of spectral clustering in stochastic block models. The Annals of Statistics 43:215–237.

MARIAC, C., JEHIN, L., AND SAÏDOU A.-A.A.AND VIGOUROUX, Y. 2011. Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping. Molecular ecology 20:80–91.

MEILLEUR, B. A. 1998. Clones within clones: Cosmology and esthetics and Polynesian crop selection. Anthropologica 40:71–82.

MIKLÓS, I. AND PODANI, J. 2004. Randomization of presence-absence matrices: comments and new algorithms. Ecology 85:86–92.

MULUMBA, J. W., NANKYA, R., ADOKORACH, J., KIWUKA, C., FADDA, C., DE SANTIS, P., AND JARVIS, D. I. 2012. A risk-minimizing argument for traditional crop varietal diversity use to reduce pest and disease damage in agricultural ecosystems of Uganda. Agriculture, Ecosystems & Environment 157:70–86.

PADOCH, C. AND JONG, W. D. 1991. The house gardens of Santa Rosa: Diversity and variability in an Amazonian agricultural system. Economic Botany 45:166–175.

PERONI, N. AND HANAZAKI, N. 2002. Current and lost diversity of cultivated varieties, especially cassava, under swidden cultivation systems in the Brazilian Atlantic Forest. Agriculture, Ecosystems & Environment 92:171–183.

PERRAULT-ARCHAMBAULT, M. AND COOMES, O. T. 2008. Distribution of agrobiodiversity in home gardens along the Corrientes River, Peruvian Amazon. Economic Botany 62.

PODANI, J. AND SCHMERA, D. 2012. A comparative evaluation of pairwise nestedness measures. Ecography 35:889–900.

REYES-GARCÍA, V., MOLINA, J. L., CALVET-MIR, L., ACEITUNO-MATA, L., LASTRA, J. J., ONTILLERA, R., PARADA, M., PARDO-DE SANTAYANA, M., RIGAT, M., AND VALLÈS, J. 2013. 'Tertius gaudens': germplasm exchange networks and agroecological knowledge among home gardeners in the Iberian Peninsula. Journal of ethnobiology and ethnomedicine 9:1–11.

RIVAL, L. AND MCKEY, D. 2008. Domestication and diversity in manioc (*Manihot esculenta Crantz ssp. esculenta, Euphorbiaceae*). Current Anthropology 49:1119–1128.

RODRÍGUEZ-GIRONÉS, M. A. AND SANTAMARÍA, L. 2006. A new algorithm to calculate the nestedness temperature of presence–absence matrices. Journal of Biogeography 33:924–935.

SALICK, J., CELLINESE, N., AND KNAPP, S. 1997. Indigenous diversity of Cassava: Generation, maintenance, use and loss among the Amuesha, Peruvian upper Amazon. Economic Botany 51:6–19.

SAMBERG, L., SHENNAN, C., AND ZAVALETA, E. 2013. Farmer seed exchange and crop diversity in a changing agricultural landscape in the southern highlands of Ethiopia. Human Ecology 41:477–485.

SUBEDI, A., CHAUDHARY, P., BANIYA, B. K., RANA, R. B., TIWARI, R. K., RIJAL, D. K., STHAPIT, B. R., AND JARVIS, D. 2003. Who maintains crop genetic diversity and how?: Implications for On-farm Conservation and Utilization. Culture & Agriculture 25:41–50.

TAPIA, M. E. 2000. Mountain Agrobiodiversity in Peru. Mountain Research and Development 20:220–225.

THÉBAULT, E. AND FONTAINE, C. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science 329:853–856.

TUXILL, J., REYES, L. A., MORENO, L. L., UICAB, V. C., AND JARVIS, D. I. 2010. All maize is not equal: maize variety choices and mayan foodways in rural Yucatan, Mexico, pp. 467–486. In Pre-Columbian Foodways: interdiscipilanary approaches to food, culture, and markets in Ancient Mesoamerica. Springer, j.e. staller and m.d. carrasco edition.

ULRICH, W. AND GOTELLI, N. J. 2012. A null model algorithm for presence–absence matrices based on proportional resampling. Ecological Modelling 244:20–27.

VIGOUROUX, Y., BARNAUD, A., SCARCELLI, N., AND THUILLET, A.-C. 2011. Biodiversity, evolution and adaptation of cultivated crops. Comptes rendus biologies 334:450–457.

WASSERMAN, S. AND FAUST, K. 1994. Social network analysis: methods and applications. Cambridge University Press.

ZAMAN, A. AND SIMBERLOFF, D. 2002. Random binary matrices in biogeographical ecology—instituting a good neighbor policy. Environmental and Ecological Statistics 9:405–421.

ZIMMERER, K. S. 1991. Labor shortages and crop diversity in the Southern Peruvian Sierra. Geographical Review 81:414–432.

Supplementary Material

Statistical power study of the contribution test

The same model as in Section 4.5.4 are used for studying the behavior of the contribution test and especially their power. These different settings of parameters correspond to a global density of approximatively 0.18. 1000 incidence matrices were simulated in each of the three cases with different sizes: $n = 20, 50$ and $m = 20, 50$.

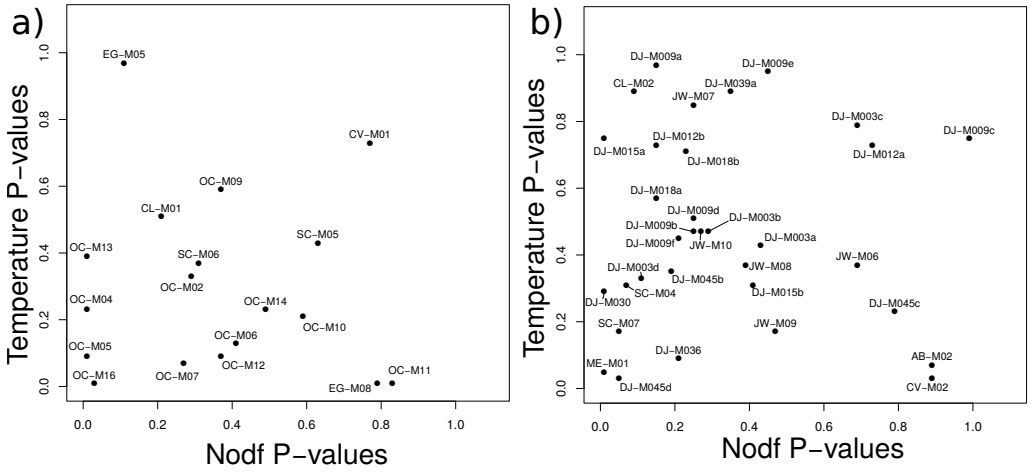

Figure 15: Plot representing in X-axis the NODF P-values computed by re-sampling, and in Y-axis the Temperature P-values computed by re-sampling. In the both cases, re-sampling was performed using the configuration model: a) for data sets collected at the specific level, b) for data sets collected at the infra-specific level.

Tables P.: ref.. present the proportion of rejection (in %) when the α level is set to 1% and 5% for the different sizes of incidence matrices. The tests have globally the same power. When there are only $n = 20$ households or $m = 20$ plants, the power is quite low. For $n = m = 50$, the power is totally satisfactory. In the null model when there is no interaction between richness of the households and the status of plants, the p-values are nearly uniformly distributed on $[0,1]$ as expected under a null model.

Estimation of nestedness

This section describes the nestedness results obtained on the meta-data set using two methods: the Temperature (Rodríguez-Gironés and Santamaría, 2006) and the NODF (Almeida-Neto et al., 2008). The figure 15 represents the P-values computed for each estimator after re-sampling using the configuration model introduce in section 4.4.1. Our results are consistent with Podani and Schmera (2012) because for the same meta-data set, tests performed with one or another index were inconsistent.