



HAL
open science

A Network-Based Method to Detect Patterns of Local Crop Biodiversity: Validation at the Species and Infra-Species Levels CINVESTAV-IPN

Mathieu Thomas, Nicolas Verzelen, Pierre Barbillon, Oliver T. Coomes, Sophie Caillon, Doyle Mckey, Marianne Elias, Eric Garine, Christine Raimond, Edmond Dounias, et al.

► **To cite this version:**

Mathieu Thomas, Nicolas Verzelen, Pierre Barbillon, Oliver T. Coomes, Sophie Caillon, et al.. A Network-Based Method to Detect Patterns of Local Crop Biodiversity: Validation at the Species and Infra-Species Levels CINVESTAV-IPN. *Advances in Ecological Research*, 2015, Ecosystem Services From Biodiversity to Society, Part 1, 10.1016/bs.aecr.2015.10.002 . hal-02799313v2

HAL Id: hal-02799313

<https://hal.science/hal-02799313v2>

Submitted on 6 Jan 2020

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

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

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 Verzelen⁴, Pierre Barbillon⁵,
Oliver T. Coomes⁶, Sophie Caillon², Doyle McKey^{2,7}, Marianne Elias⁸,
Eric Garine⁹, Christine Raimond¹⁰, Edmond Dounias², Devra Jarvis^{11,12}, Jean Wencélius⁹,
Christian Leclerc¹³, Vanesse Labeyrie¹³, Pham Hung Cuong¹⁴, Nguyen Thi Ngoc Hue¹⁴,
Bhuwon Sthapit¹⁵, Ram Bahadur Rana¹⁶, Adeline Barnaud^{17,18,19},
Chloé Violon⁹, Luis Manuel Arias Reyes²⁰, Luis Latournerie Moreno²¹,
Paola De Santis²², François Massol²³

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 Bioersivity 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 Posgrado e Investigación 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

1
2 In this paper we develop new indicators and statistical tests to characterize patterns of crop
3 diversity at local scales. Households growing a large number of species or landraces are known
4 to contribute an important share of local available diversity of both rare and common plants
5 but the role of households with low diversity remain little understood: do they grow only com-
6 mon varieties – following a nestedness pattern typical of mutualistic networks in ecology – or
7 do ‘diversity poor’ households also grow rare varieties? This question is pivotal in ongoing ef-
8 forts to assess the contribution of small farmers to global agrobiodiversity at local scales. We
9 develop new network-based approaches to characterize the distribution of local crop diversity
10 at the village level (species and infra-species) and validate these approaches using meta-data
11 sets from 10 countries. Our results highlight the sources of heterogeneity in the local crop di-
12 versity. We often identify two or more groups of households based on their different levels of
13 diversity. In some datasets, ‘diversity poor’ households significantly contribute to the local crop
14 diversity. Generally, we find that the distribution of crop diversity is more heterogeneous at the
15 species than at the infra-species level. This analysis reveals the absence of a general pattern of
16 crop diversity distribution independent of agro-ecological and socio-cultural context.

17 **Contents**

18 **1 Glossary** **6**

19 **2 Introduction** **7**

20 **3 Description of the datasets used in the meta-analysis** **11**

21 **4 Description of the methodological framework** **13**

22 4.1 Mathematical formalism 13

23 4.2 Variability of households' and plants' degrees 15

24 4.2.1 Description of the test on degree distributions 15

25 4.2.2 Application of the test on degree distributions to a toy example 16

26 4.3 Revealing data structure through latent block models 18

27 4.3.1 Description of the latent block models 18

28 4.3.2 Application of LBM to a toy example 19

29 4.4 Uncovering outliers through principal component analysis 21

30 4.4.1 Configuration model 21

31 4.4.2 Principal Component analysis (PCA) on residuals 22

32 4.4.3 Goodness-of-fit test of the configuration model 23

33 4.4.4 A new representation of the incidence matrix 23

34 4.4.5 Toy-examples 23

35 4.5 Measuring originality of households' contributions through diversity measures . . . 28

36 4.5.1 Theoretical framework 28

37 4.5.2 Measuring the diversity cultivated by plant-poor and plant-rich households . 29

38 4.5.3 Measuring the impact of plant-poor and plant-rich households 30

39 4.5.4 Measuring originality of households' contributions through diversity measures

40 on toy examples 31

41 **5 Patterns of local crop diversity: results of the meta-analysis** **34**

42 5.1 Variability of households' and plants' degrees 34

43	5.2	Structure detection through model-based clustering (LBM)	35
44	5.3	Outlier detection through PCA	36
45	5.4	Households' contributions to local diversity	37
46	6	Discussion	38
47	6.1	Contrasted patterns of local crop diversity at the species and infra-species levels . .	38
48	6.2	Relevance of the network-based methods	42
49	7	Conclusion	44

50 1 Glossary

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

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

53 **Edge:** an edge is a link between two vertices, every edge has two endpoints in the sets of vertices.

54 In the particular case of bipartite networks, the two endpoints belong to two disjoint subsets of
 55 nodes, *e.g.* farmer households (H) and crops (S , species or landraces). The presence of an edge
 56 indicates that the household grows the considered crop.

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

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

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

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

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

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

73 **Latent block models:** random graph models assuming a mixture distribution both on rows (house-
 74 holds) and columns (plants). Households and plants are assumed to belong to blocks which are

75 latent (not observed). Thanks to a statistical inference procedure this block structure can be un-
 76 covered (see Section 4.3).

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

80 2 Introduction

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

91 A primary requisite to understanding and predicting the sustainability of agricultural systems
 92 facing environmental, political, social and economical changes is to assess how these systems can
 93 maintain crop diversity (*e.g.* Samberg et al., 2013). For instance, in the case of manioc managed
 94 by Makushi Amerindians of Guyana, some varieties are specially grown for special dishes, others
 95 for another use; some grow quickly, thereby ensuring early yield, while others grow slowly and act
 96 ever-present insurance (Elias et al., 2000). Often, diversity is just valued for its own sake (Boster,
 97 1985), or as a means to foster social relations (Heckler and Zent, 2008). Another example is the
 98 great diversity of landraces present in milpas of Yucatan, the end product of several thousand years
 99 of directed selection on maize, beans, squash and chile peppers by the region's farmers. Under-
 100 standing landraces relationships, it is possible to gain insight into the cultural history of crops in
 101 Yucatan. The particular traits exhibited by local varieties grown in milpas today reflect Yucate-

102 can farmers' short- and long-term responses to agroenvironmental conditions, the ecological de-
103 mands of crop production, and the aesthetic, culinary, and religious sensibilities of farmers (Tuxill
104 et al., 2010). Maintaining crop diversity is of paramount importance in helping crops and farmers
105 adapt to global changes, notably climate change (Vigouroux et al., 2011) and the increasingly rapid
106 emergence of agricultural pests (Diamond, 2002). In the face of such change, drastically reduced
107 diversity of crop species and varieties would inevitably lead to increasingly unpredictable yields.
108 In addition, cultivating diverse crops and varieties at the landscape level favors ecological and eco-
109 nomic sustainability by reducing the need for chemical inputs (Bianchi et al., 2006; Crowder et al.,
110 2010).

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

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

131 Coppens d'Eeckenbrugge, 2012).

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

144 Our main goal in this paper is to answer the question “which households contribute, and how,
145 to the diversity of crops grown in a given village?” by examining on inventories of crops species and
146 landraces grown at the household level. To do so, we offer a novel methodological framework us-
147 ing network-based and null model-based statistical tests. From a methodological perspective, in-
148 ventory datasets can be construed as bipartite networks*, namely crop-by-household interaction
149 networks, in the same way as plant-pollinator or host-parasite interaction networks in ecology. In
150 social network analysis, network approaches have been used to assess the properties of network
151 processes linked to social institutions such as friendship, advice or seed exchange networks (“who
152 interacts with whom” or “who gives to whom”) (Wasserman and Faust, 1994; Lazega et al., 2012;
153 Reyes-García et al., 2013). In ecology, on the other hand, networks have been used to study both
154 contact networks (metapopulations or metacommunities) and structured interaction networks*
155 such as food webs (*e.g.* host-plant networks) or mutualistic networks (*e.g.* plant-pollinator net-
156 works). When interaction partners can be clearly categorized (*e.g.* plants *vs.* pollinators; plants,
157 herbivores and parasitoids), the use of bi- or multi-partite networks is an appropriate approach. In
158 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

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

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

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

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

206 **3 Description of the datasets used in the meta-analysis**

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

Patterns of local crop biodiversity

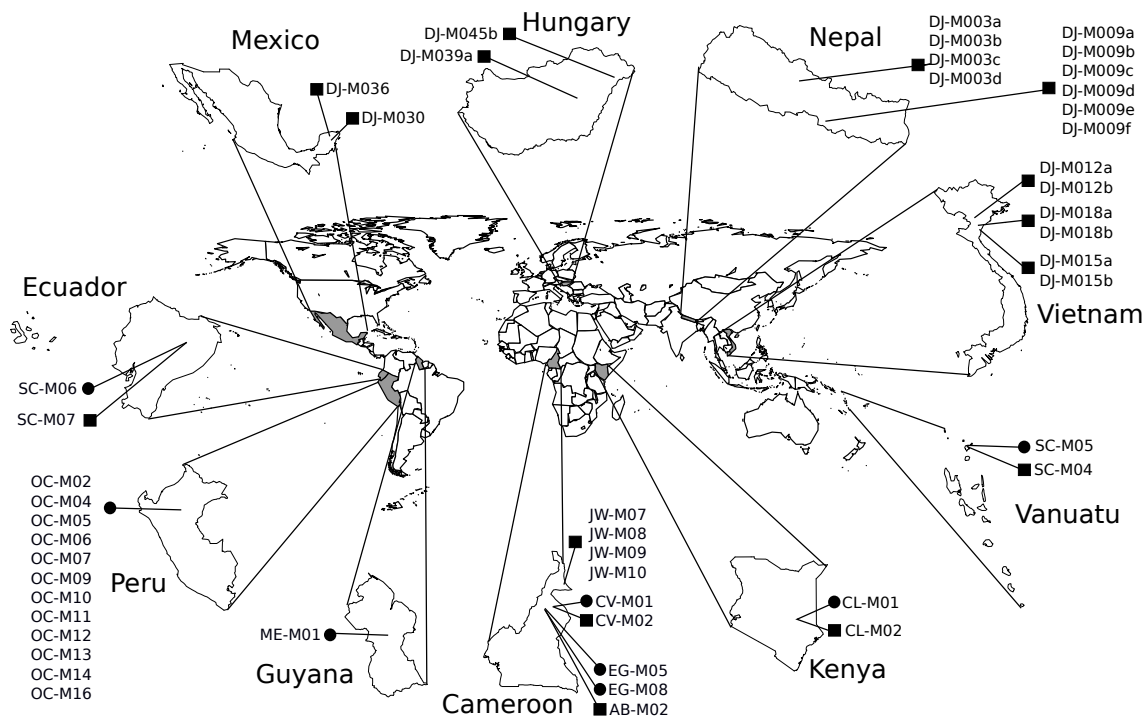


Figure 1: Map locating the different data sets used in the meta-analysis. Filled circles correspond to the data sets collected at the specific level and filled squares correspond to the data set collected at the infra-specific level

215 for 32 datasets, information was collected at the landrace level, which corresponds to the termi-
 216 nal taxon in the farmer local naming systems, covering seven different species (maize, rice, wheat,
 217 bean, manioc, taro and sorghum) which correspond to the major crops of the under area (Table
 218 2). These species are characterized by their predominant propagation mode with partially out-
 219 crossing, outcrossing, inbreeding and clonal following the classification proposed by Jarvis et al.
 220 (2008). Data were structured following a rectangular incidence matrix* with households in rows
 221 and species or landraces in columns, and represented as a bipartite network. Data collected at the
 222 species or infra-species level represent two levels of local crop biodiversity. Underlying processes
 223 shaping the distribution of local crop diversity are assumed to be different for these two levels.
 224 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.

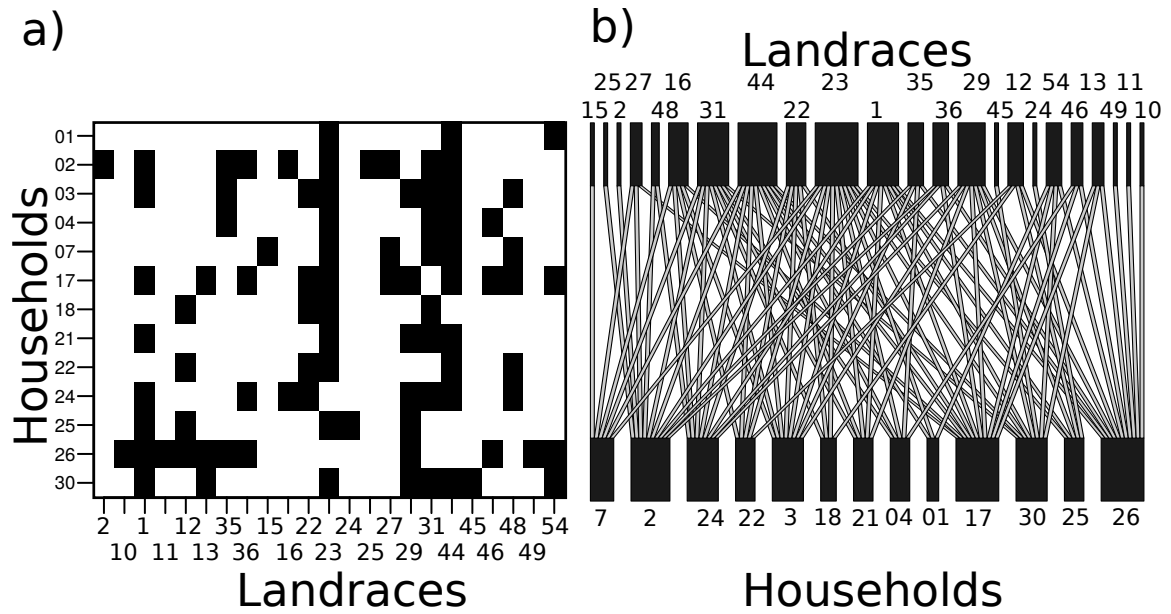


Figure 2: a) Example of incidence matrix where 0 are black cells and 1 are white cells; b) example of bipartite network between households and landraces (dataset AB-M02).

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

$$S_i = \sum_j X_{ij}. \quad (1)$$

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

$$F_j = \sum_i X_{ij}. \quad (2)$$

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

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

255 Following the network terminology, S_i is also called the household's degree and F_j the plant's
 256 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 $\widehat{\text{Var}}(S) = \frac{1}{n-1} \sum_{i=1}^n (S_i - m\hat{p})^2$ is the empirical variance of (S_i) , $i = 1, \dots, n$. 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 $i = 1, \dots, n_{sim}$, 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

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

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

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

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

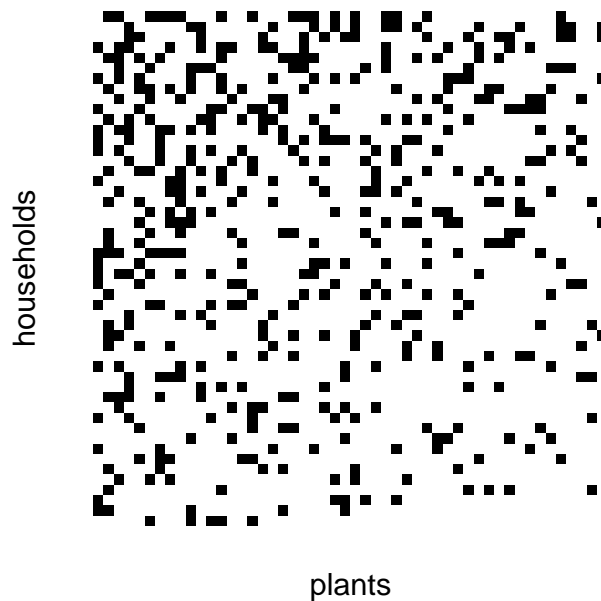


Figure 3: Incidence matrix with entries generated independently and identically distributed according to a Bernoulli distribution with probability 0.2.

276 Under a null model, called Erdős-Renyii, where all the entries of X follow independent Bernoulli
 277 distribution with identical parameter, the households' degrees and the plants' degrees follow bino-
 278 mial distributions. consequently, any small p -value ($pval_{L,row}$, $pval_{R,row}$, $pval_{L,col}$, $pval_{R,col}$) entail
 279 that this Erdős-Renyii model is not realistic.

280 4.2.2 Application of the test on degree distributions to a toy example

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

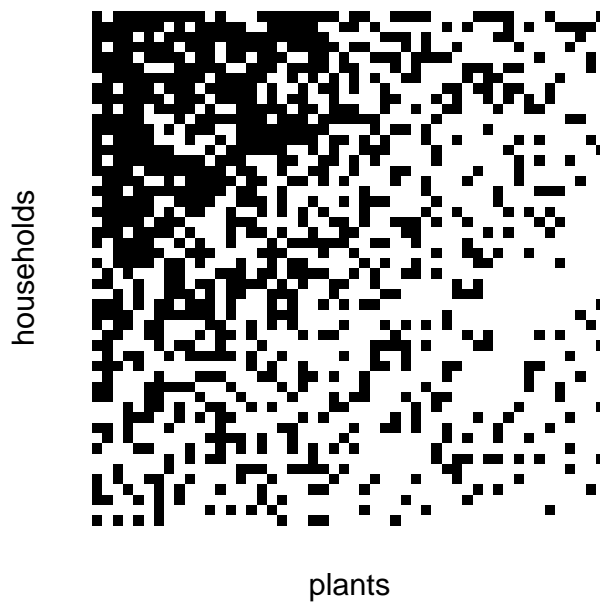


Figure 4: Incidence matrix generated with heterogeneous distribution for different groups of plants and households (see Figure 7 in next subsection for details). Some households grow more plants than other and some plants are more popular.

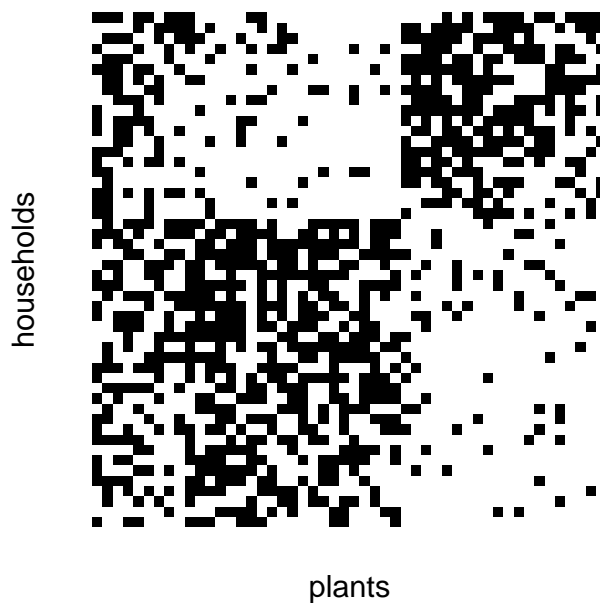


Figure 5: Incidence matrix generated with distribution implying particular association between plants and households (see Figure 6 in next subsection for details). Two groups of plants are mainly grown by corresponding subgroups of households.

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

294

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

299 **4.3 Revealing data structure through latent block models**

300 **4.3.1 Description of the latent block models**

301 In order to cluster the households and the plants simultaneously on the basis of the incidence
 302 matrix \mathbf{X} , we propose to use a probabilistic model called Latent Block Model (Govaert and Nadif,
 303 2008; Keribin et al., 2014). It consists in assuming a mixture distribution both on the households
 304 and on the plants. According to this model, the network is generated according to latent blocks
 305 (also called clusters) of household and latent block of households. Conditioned to these latent
 306 blocks, the probability that a household i grows a plant j only depends on the block $V(i)$ to which
 307 household i belongs and the block W_j to which plant j belongs. For all $1 \leq i \leq n$, $1 \leq j \leq m$,
 308 $1 \leq q \leq Q$, $1 \leq l \leq L$, the probability that i belongs to block q , that j belongs to block l and the
 309 conditional probability of X_{ij} given the block V_i and W_j are respectively denoted

$$\mathbb{P}(V_i = q) = \alpha_q,$$

$$\mathbb{P}(W_j = l) = \beta_l,$$

$$\mathbb{P}(X_{ij} = 1 | V_i = q, W_j = l) = \pi_{ql},$$

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

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

316
 317 The standard procedures to obtain maximum likelihood estimated when dealing with latent
 318 variables rely on the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). However,
 319 the computation of the conditional distribution of the latent variables with respect to the observed
 320 data is not tractable which makes the E step infeasible. Following Govaert and Nadif (2008), we
 321 use a variational approach to cope with this difficulty. The number of blocks of households Q and
 322 the number of blocks of plants L are chosen thanks to the integrated completed likelihood (ICL)
 323 criterion as proposed in Keribin et al. (2014). Once the parameters have been estimated, we obtain
 324 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
 325 are estimated. We can then provide a new representation of the incidence matrix \mathbf{X} where the rows
 326 (households) and the columns (plants) have been reorganized in homogeneous blocks. We used
 327 the R package (Leger, 2015) to perform the estimations and the model selection.

328 **4.3.2 Application of LBM to a toy example**

329 Figures 6, 7 and 8 are illustrations of the block clustering provided by the LBM in three typical
 330 cases. The cases of Figure 6 and 7 are the same as those in Figures 5 and 4 respectively. The
 331 groups were considered as latent/unknown and the households and plants were clustered in ho-
 332 mogeneous blocks by using the inference procedure described above. This is illustrated in Figure
 333 6 where the same incidence matrix is plotted before and after re-organization according to the es-
 334 timated blocks. In Figure 6, the difference between the two groups of households comes from the
 335 two last groups of plants. The first group of plants is equally grown up by households of any group.
 336 On the contrary, the second group of plants is mainly grown up by the second group of households
 337 and the third group of plants is mainly grown up by the first group of households. In Figure 7, the
 338 households can be separated on the basis on the number of plants that they grown up, a group
 339 can be said to be rich and the other to be poor. Similarly, two groups are also found for plants, one

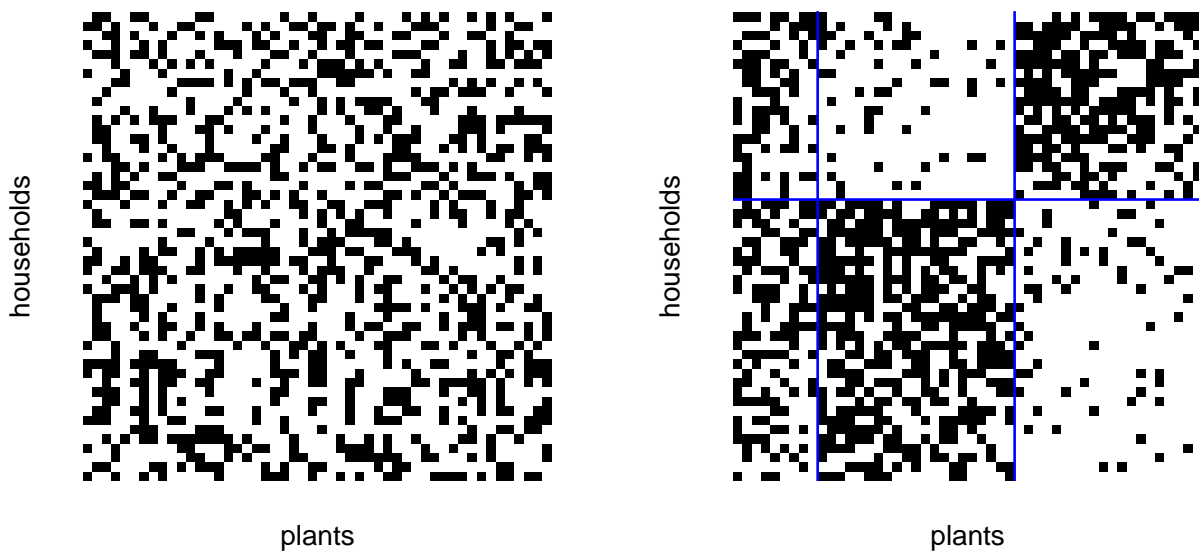


Figure 6: Incidence matrix generated according to a LBM with 3 blocks of plants, 2 blocks of households and $\pi = \begin{pmatrix} 0.5 & 0.1 & 0.6 \\ 0.5 & 0.6 & 0.1 \end{pmatrix}$. Left: observed incidence matrix. Right: same incidence matrix re-organized and clustered in homogeneous blocks obtained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

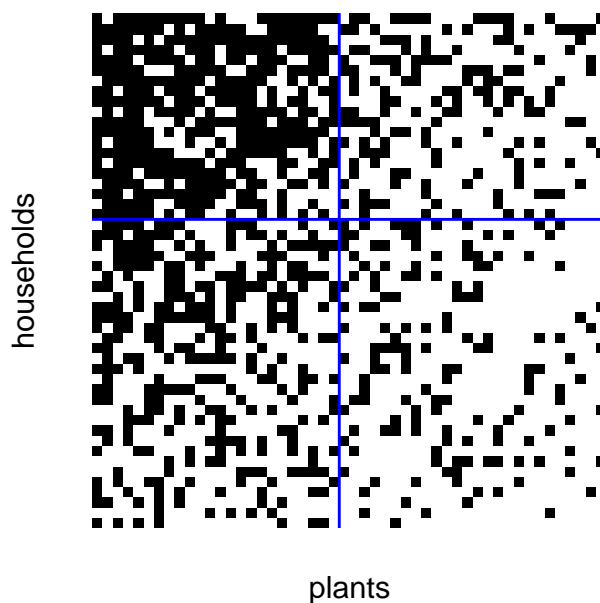


Figure 7: LBM clustering when the data are generated with 2 blocks of households (rich and poor households), 2 blocks of plants (rare and frequent plants) and $\pi = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.2 \end{pmatrix}$

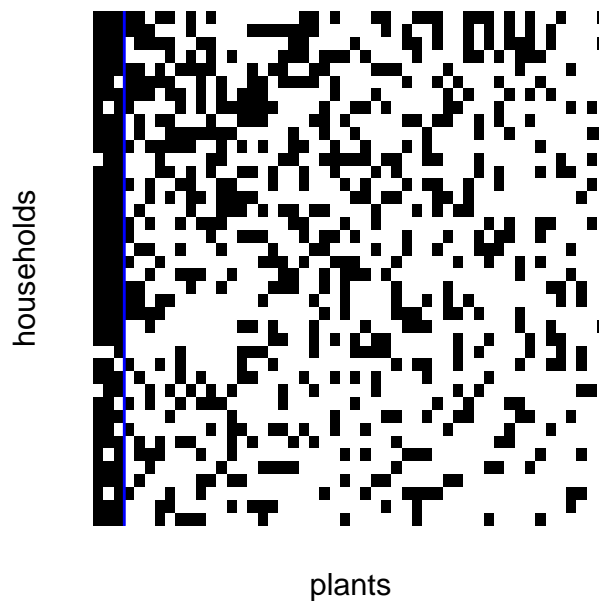


Figure 8: LBM clustering when the data are generated with 1 block of households, 2 blocks of plants (one block with only 3 plants) and $\pi = (0.9 \ 0.3)$

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

346 4.4 Uncovering outliers through principal component analysis

347 4.4.1 Configuration model

348 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) configu-
 349 ration model with parameters (S_i) and F_j is the uniform distribution over all incidence matrices
 350 that leave the degrees S_i and F_j unchanged. In the ecological literature, this model is sometimes
 351 referred as the Fixed-Fixed null model (Ulrich and Gotelli, 2012; Connor and Simberloff, 1979; Za-
 352 man and Simberloff, 2002). In contrast to the LBM, the configuration model takes for given that
 353 some households might grow much more plants than others and that some plants are more com-

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

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

361 4.4.2 Principal Component analysis (PCA) on residuals

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

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

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

377 4.4.3 Goodness-of-fit test of the configuration model

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

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

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

391 4.4.4 A new representation of the incidence matrix

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

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

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

397 4.4.5 Toy-examples

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

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

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

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

421 In the last example, we draw random variables a_i and b_j as above. Then, the households
 422 are divided in two groups A_1 and A_2 of size $n/2$ and the plants are divided in two groups B_1
 423 and B_2 of size $m/2$. Then, the entry X_{ij} is drawn according to Bernoulli distribution with pa-
 424 rameter $\min(p_{in} 2a_i b_j, 1)$ 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
 425 $(i, j) \in A_1 \times B_2$ or $(i, j) \in A_2 \times B_1$ with $p_{in} = 1.4$ and $p_{out} = 0.6$. Intuitively, the households from A_1
 426 (resp. A_2) preferentially grow plants from B_1 (resp. B_2), but the model also allows the degree of the
 427 household and each plant to be heterogeneous inside the blocks. As a consequence, this model,
 428 called degree-corrected is neither a LBM with 2×2 blocks nor a configuration models but a blend

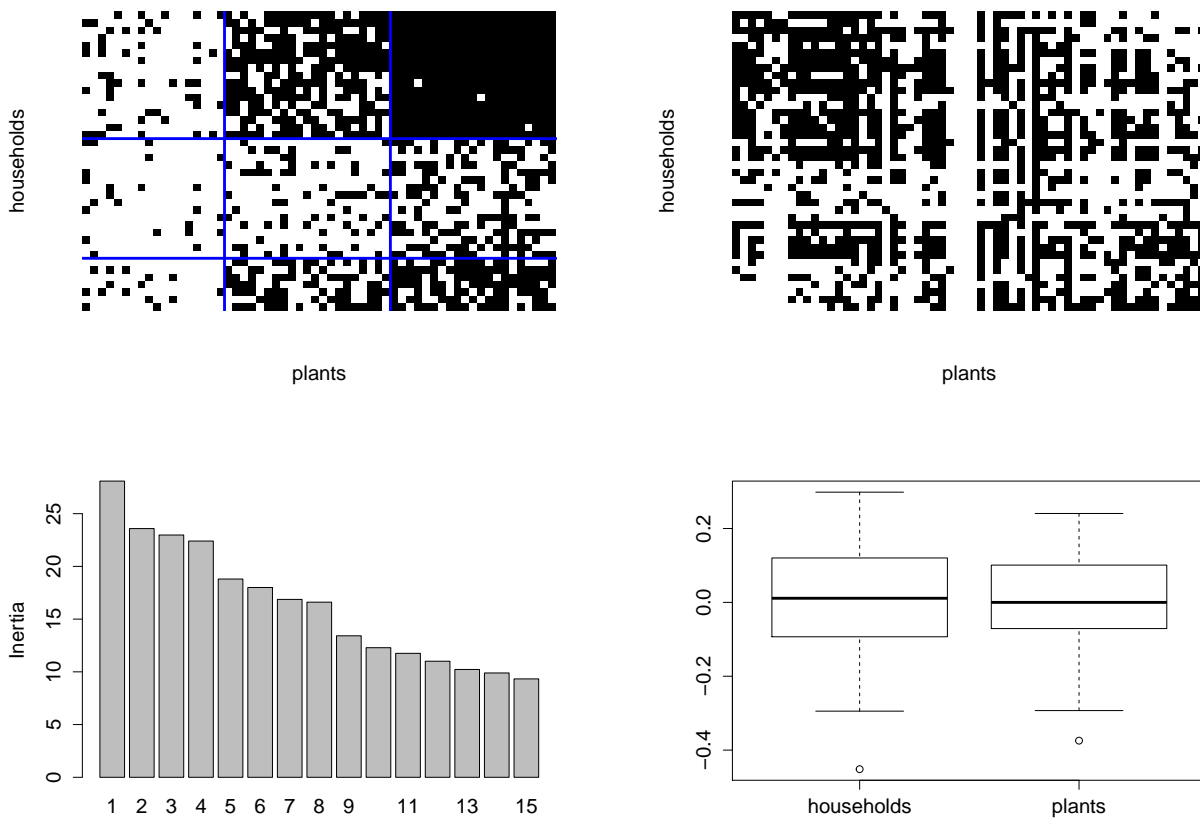


Figure 9: First example. The upper-left panel is the LBM representation. The lower left panel is the scree plot of the residuals PCA. The upper left panel is the representation of the incidence matrix according to the PCA ordering (5). The boxplots of the PCA first coordinates are pictured in the lower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

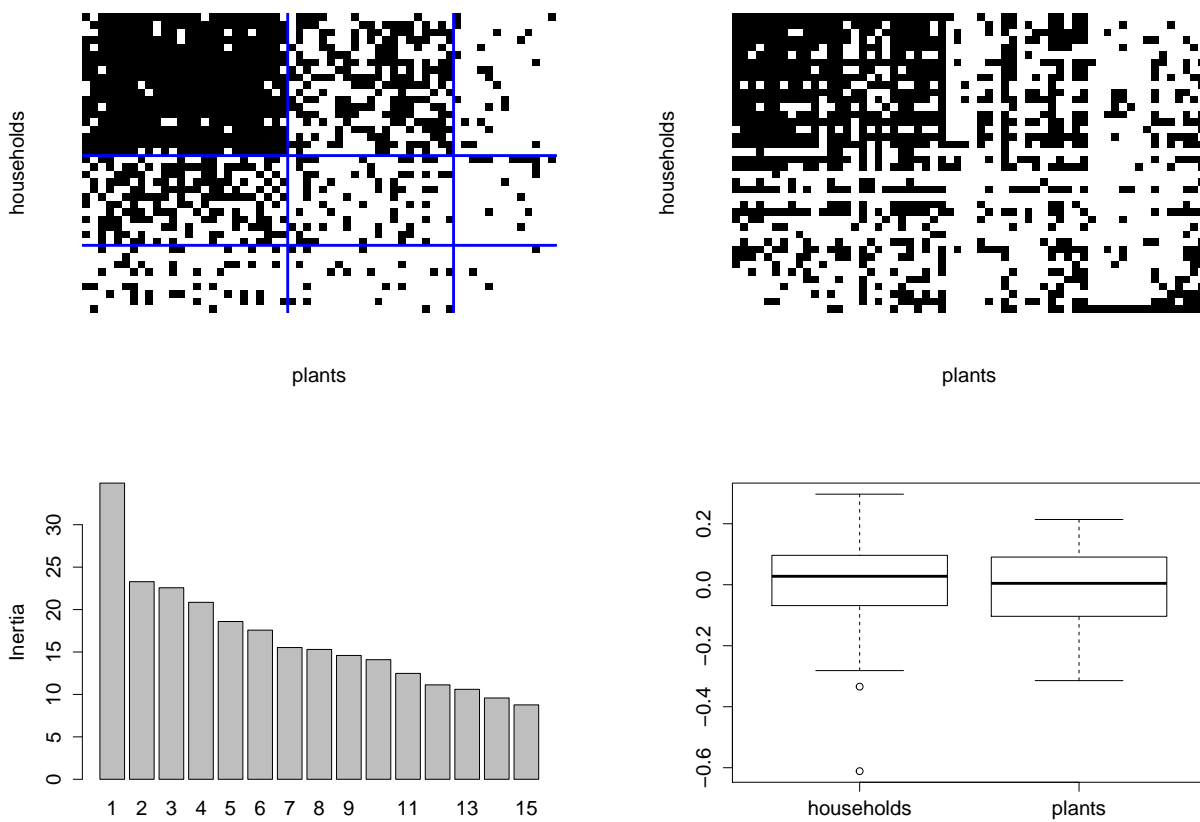


Figure 10: Second example. The upper-left panel is the LBM representation. The lower left panel is the scree plot of the residuals PCA. The upper left panel is the representation of the incidence matrix according to the PCA ordering (5). The boxplots of the PCA first coordinates are pictured in the lower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

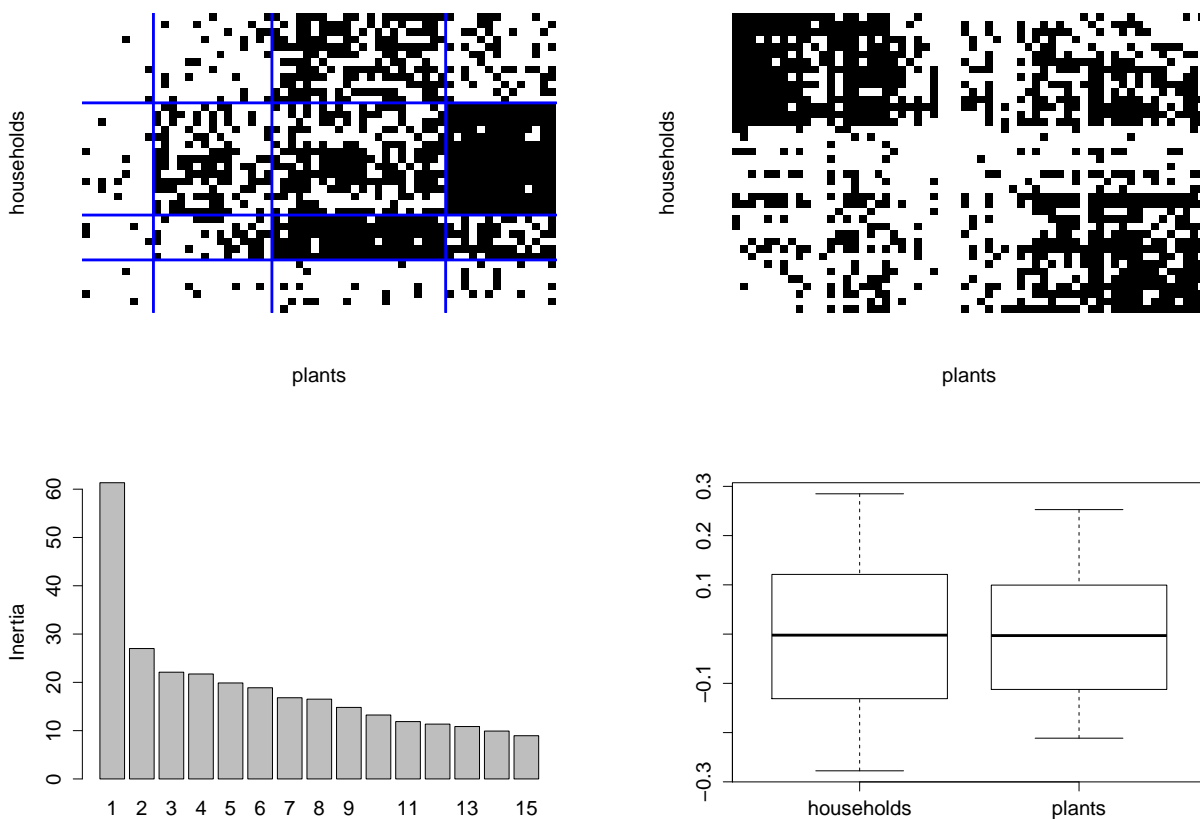


Figure 11: Third example. The upper-left panel is the LBM representation. The lower left panel is the scree plot of the residuals PCA. The upper left panel is the representation of the incidence matrix according to the PCA ordering (5). The boxplots of the PCA first coordinates are pictured in the lower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

429 of them. The LBM estimation method recovers too many blocks (Figure 11) by grouping farms or
 430 plants being in the same group and having similar degrees. The p -value for the configuration test
 431 is found to be smaller than 10^{-3} (see also the scree plot). Contrary to the previous example, this
 432 unusually large singular values is not due to outliers (see lower-right panel) but to the presence
 433 of a block structure. The PCA-based matrix representation highlights the presence of these two
 434 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 :

$$p_{ij} = \frac{X_{ij}}{S_i} \quad (6)$$

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 :

$$q_i = \frac{S_i}{N} \quad (7)$$

$$h_j = \frac{F_j}{N} \quad (8)$$

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

$$H_i = - \sum_j p_{ij} \log p_{ij} = \log S_i \quad (9)$$

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

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

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

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

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

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

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

$$H_\beta = \sum_i q_i H_{iT} \quad (13)$$

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

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

462 4.5.2 Measuring the diversity cultivated by plant-poor and plant-rich households

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

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

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

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

475 4.5.3 Measuring the impact of plant-poor and plant-rich households

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

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

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

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

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

$$\sigma_I = \sum_{i \in I} S_i \quad (18)$$

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

$$\varphi_{j,I} = \sum_{i \in I} X_{ij} \quad (19)$$

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

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

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

487 Again, we assess the significance of $H_{\beta,I}$ by a permutation test based on the configuration
 488 model. As the set I contains all households belonging to a certain quantile of the distribution
 489 of S_i , all realizations of the incidence matrix \mathbf{X} under the configuration model preserve the set of
 490 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
 491 equation 16 is invariant with respect to the configuration model. The quantity D_I in the right-hand
 492 side of equation 16, however, does not satisfy this invariance. Thus, large values of $H_{\beta,I}$ unusually
 493 large for the configuration model mean that households in subset I contribute more to cultivated
 494 biodiversity than expected by the number of types cultivated by households in I .

495 4.5.4 Measuring originality of households' contributions through diversity measures on toy 496 examples

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

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

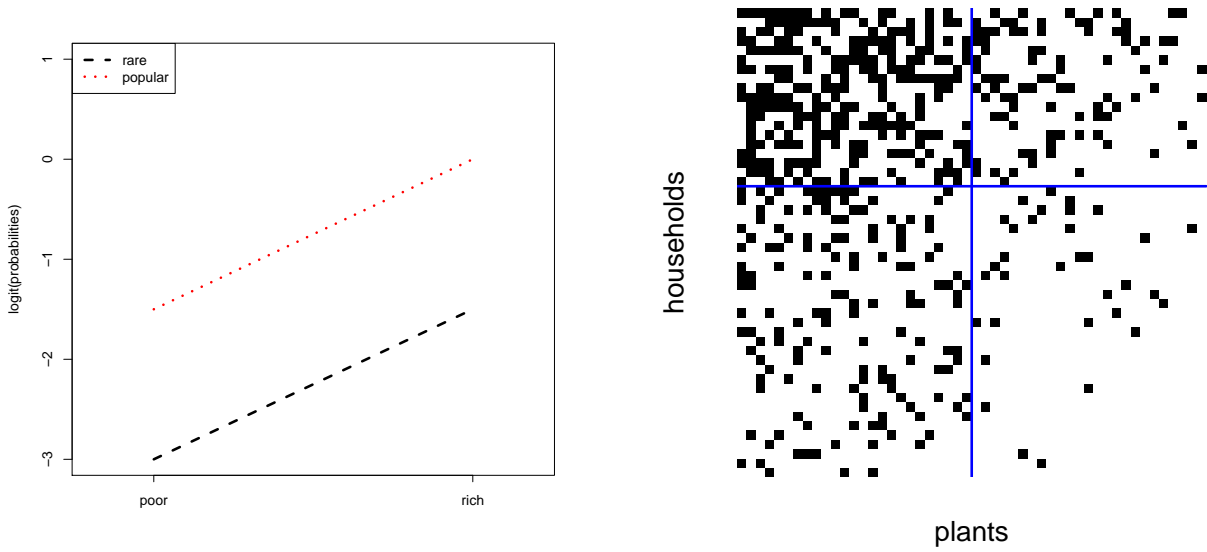


Figure 12: Toy example with equal contribution to diversity for rich and poor households. $\mu = -3$, $\alpha(\text{rich}) = \beta(\text{popular}) = 1.5$, $\gamma(\text{rich}, \text{popular}) = 0$. Left: probabilities for a household to grow a plant. Right: Incidence matrix.

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

$$\begin{aligned} \alpha(\text{poor}) &= \beta(\text{rare}) = \gamma(\text{poor}, \text{rare}) = \gamma(\text{rich}, \text{rare}) = \gamma(\text{poor}, \text{popular}) = 0 \\ \alpha(\text{rich}) &> 0, \beta(\text{popular}) > 0, \gamma(\text{rich}, \text{popular}) \neq 0 \\ \alpha(\text{rich}) + 0.5 \cdot \gamma(\text{rich}, \text{popular}) &> \alpha(\text{poor}) = 0 \\ \beta(\text{popular}) + 0.40 \cdot \gamma(\text{rich}, \text{popular}) &> \beta(\text{rare}) = 0 \end{aligned}$$

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

507 **Three contrasted toy examples:** Figures 12, 13 and 14 correspond respectively to the three fol-
 508 lowing cases:

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

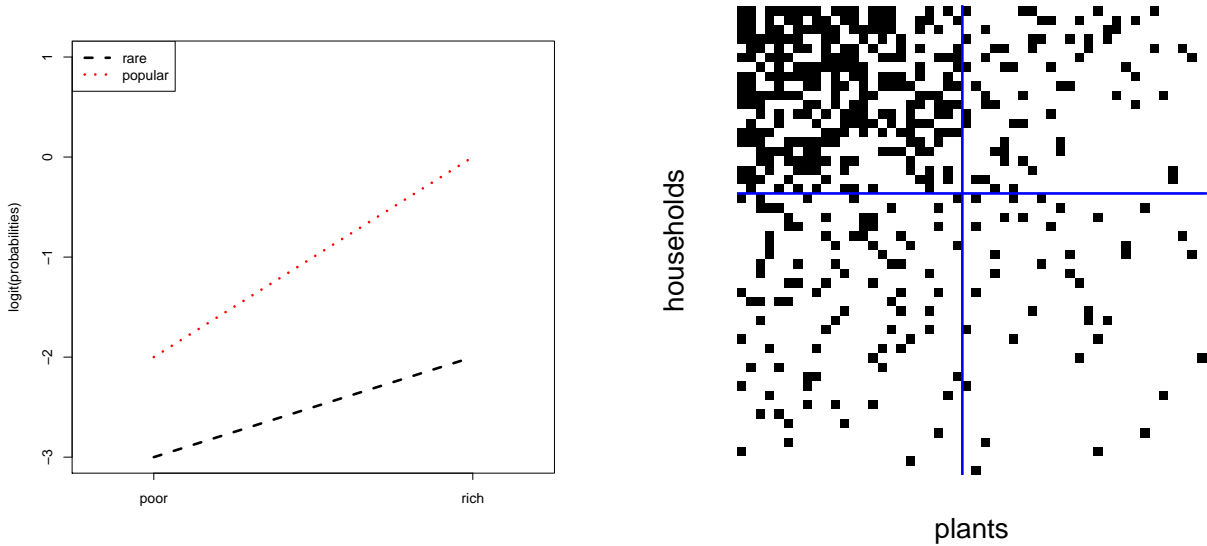


Figure 13: Toy example with greater contribution to diversity of poor households. $\mu = -3$, $\alpha(\text{rich}) = \beta(\text{popular}) = \gamma(\text{rich}, \text{popular}) = 1$. Left: probabilities for a household to grow a plant. Right: Incidence matrix.

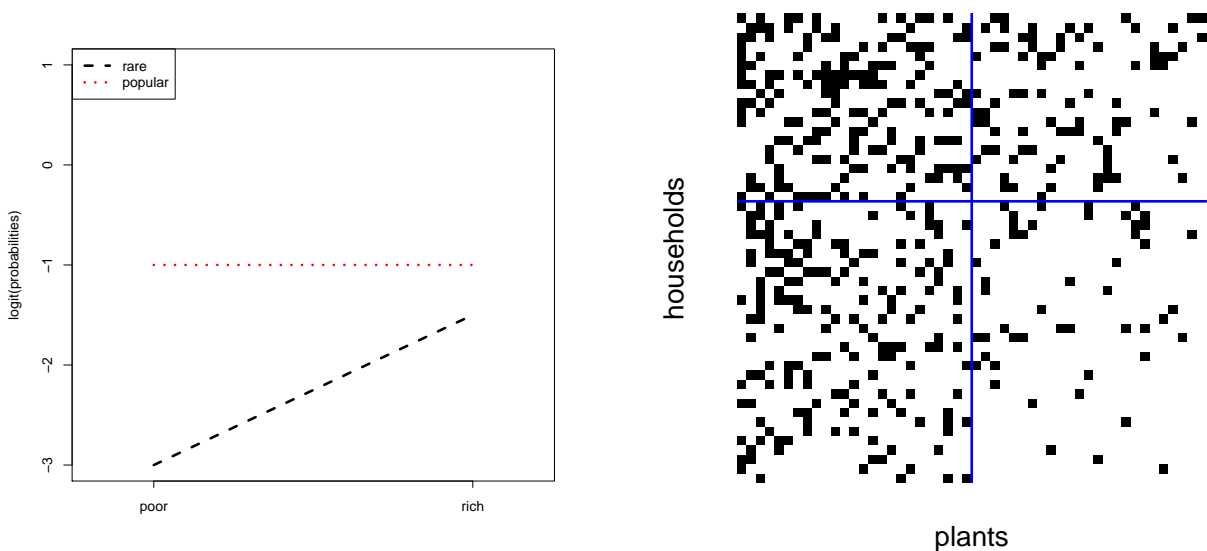


Figure 14: Toy example with greater contribution to diversity of rich households. $\mu = -3$, $\alpha(\text{rich}) = -\gamma(\text{rich}, \text{popular}) = 1.5$, $\beta(\text{popular}) = 2$. Left: probabilities for a household to grow a plant. Right: Incidence matrix.

511 fixed to 0.

- 512 2. The poor households have a greater contribution to diversity since they grow with nearly the
513 same probability rare plants and popular plants while this probability of growing rare plants
514 is clearly smaller than the probability of growing popular plants for rich households.
- 515 3. The rich households have a greater contribution to diversity. They are richer since they have
516 the same ability of growing rare and popular plants.

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

524 **5 Patterns of local crop diversity: results of the meta-analysis**

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

527 **5.1 Variability of households' and plants' degrees**

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

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

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

547 **5.2 Structure detection through model-based clustering (LBM)**

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

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

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

569 **5.3 Outlier detection through PCA**

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

572 **Species diversity** Using the test introduced in section 4.4.3, H_0 was rejected 9 times over the
 573 18 datasets at the $\alpha = 0.05$; rejecting H_0 highlights the existence of outliers. These outliers are
 574 generally two or three per dataset and can be characterized as households that grow a different
 575 subset of species compared to other households with an equivalent degree, *i.e.* belonging to the
 576 same cluster.

577 **Infra-species diversity** H_0 was rejected for only four datasets over the 32 datasets (CL-M02, DJ-
 578 M018a, DJ-M018b, DJ-M030). These results indicate that in addition to growing almost the same
 579 number of landraces, households from the same village grow globally the same portfolio of lan-
 580 draces. Note that for these four datasets, only one cluster was detected with the LBM (CL-M02,
 581 DJ-M018a, DJ-M018b, DJ-M030). Therefore, in this case we have households growing a particular
 582 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 (Gariné 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 (Gariné 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 staples 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 portfolio composition. Our analysis of
 691 outliers identified certain households holding unique portfolios 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 nestedness and of a sink-source dynamics described in

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

698 Nevertheless, it would be incorrect to say that it is a consistent tendency in the meta-data set.
699 Indeed, we observed the opposite relationship in other data sets whereby low diversity households
700 contributed significantly to the local diversity (EG-M05, EG-M08, DJ-M015a, SC-M07). In some
701 case, one or a few farmers grow rarer species or landraces owing to curiosity, for aesthetic reasons,
702 or to maintain a social status of expert at the local level (Elias et al., 2000; Meilleur, 1998; Hawkes,
703 1983), or to have an object that the others do not have (Coomes and Ban, 2004). Possessing an in-
704 frequent species or landrace might, for instance, allow a young farmer to distinguish himself from
705 others in societies independently of economic capital. Having an object that others do not have,
706 could increase the value of the eventual transfer to other members of the community. Cultivating
707 rarity helps both to gain a social status within the village and to have highly valued objects to ex-
708 change (Caillon and Lanouguère-Bruneau, 2005). Additional factors influence the distribution of
709 local crop diversity, for instance, the role played by differences associated with gender and genera-
710 tion, access to local seed markets, farmers' food preferences, and the market value of crops. Verti-
711 cal transmission of seeds and knowledge occurs between mothers-in-law and daughters-in-law in
712 patrilinear societies with virilocal residence rules, and constituting another source of divergence
713 in crop diversity between families in the same neighbourhood (Labeyrie et al., 2013).

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

720 **6.2 Relevance of the network-based methods**

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

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

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

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

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

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

759 The approach used in this paper does not rely on a direct estimation of nestedness because
 760 the different methods available to compute nestedness do not converge (Supplementary material
 761 Fig 15). However, the set of methods designed here to uncover the uniqueness of contributions to
 762 diversity of plant-rich and plant-poor households actually provide complementary information
 763 on whether specialists interact preferentially with generalists or not, as assumed under a “nested”
 764 scenario in ecology. We thus suggest that this toolkit could be used as an alternative to the clas-
 765 sic methods of nestedness detection usually applied to ecological datasets (Podani and Schmera,
 766 2012).

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

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

782 **7 Conclusion**

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

796 **Acknowledgements**

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

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

808 **References**

- 809 ABIZAID, C., COOMES, O. T., TAKASAKI, Y., AND BRISSON, S. 2015. Social Network Analysis of
 810 Peasant Agriculture: Cooperative Labor as Gendered Relational Networks. The Professional
 811 Geographer 67:447–463.
- 812 ALMEIDA-NETO, M., GUIMARÃES, P., GUIMARÃES, P. R., LOYOLA, R. D., AND ULRICH, W. 2008. A
 813 consistent metric for nestedness analysis in ecological systems: reconciling concept and mea-
 814 surement. Oikos 117:1227–1239.
- 815 ALVAREZ, N., GARINE, E., KHASAH, C., DOUNIAS, E., HOSSAERT-MCKEY, M., AND MCKEY, D.
 816 2005. Farmers' practices, metapopulation dynamics, and conservation of agricultural biodiver-
 817 sity on-farm: a case study of sorghum among the Duupa in sub-sahelian Cameroon. Biological
 818 Conservation 121:533–543.
- 819 BELLON, M. R. AND TAYLOR, J. E. 1993. "Folk" soil taxonomy and the partial adoption of new seed
 820 varieties. Economic Development and Cultural Change 41:763–786.
- 821 BIANCHI, F., BOOIJ, C. J. H., AND TSCHARNTKE, T. 2006. Sustainable pest regulation in agricul-
 822 tural landscapes: a review on landscape composition, biodiversity and natural pest control.
 823 Proceedings of the Royal Society B: Biological Sciences 273:1715–1727.
- 824 BICKEL, P. J. AND SARKAR, P. 2013. Hypothesis testing for automated community detection in
 825 networks. arXiv preprint arXiv:1311.2694 .
- 826 BISHT, I., MEHTA, P., AND BHANDARI, D. 2007. Traditional crop diversity and its conservation on-
 827 farm for sustainable agricultural production in Kumaon Himalaya of Uttaranchal State: A case
 828 study. Genetic Resources and Crop Evolution 54:345–357.
- 829 BOSTER, J. 1983. A comparison of the diversity of Jivaroan Gardens with that of the tropical forest.
 830 Human Ecology 11:47–68.
- 831 BOSTER, J. S. 1985. Selection for perceptual distinctiveness: Evidence from Aguaruna cultivars of
 832 *Manihot esculenta*. Economic Botany 39.
- 833 BRUSH, S. B., BELLON, M. R., HIJMANS, R. J., RAMIREZ, Q. O., PERALES, H. R., AND ETTEN,
 834 J. V. 2015. Assessing maize genetic erosion. Proceedings of the National Academy of Sciences
 835 112:E1–E1.
- 836 BRUSH, S. B. AND MENG, E. 1998. Farmers' valuation and conservation of crop genetic resources.
 837 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 DÉLETRE, 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.

- 875 GARINE, E. AND RAIMOND, C. 2005. La culture intensive fait-elle disparaître la biodiversité ? In
876 Dynamique de la biodiversité et modalité d'accès aux milieux et aux ressources, pp. 25–28, Paris,
877 France. Institut Français de la Biodiversité.
- 878 GAUCHAN, D., SMALE, M., AND CHAUDHARY, P. 2005. Market-based incentives for conserving
879 diversity on farms: the case of rice landraces in Central Tarai, Nepal. Genetic Resources and
880 Crop Evolution 52:293–303.
- 881 GOVAERT, G. AND NADIE, M. 2008. Block clustering with bernoulli mixture models: Comparison of
882 different approaches. Computational Statistics & Data Analysis 52:3233–3245.
- 883 HAWKES, J. G. 1983. The diversity of crop plants. p. 184pp.
- 884 HECKLER, S. AND ZENT, S. 2008. Piaroa manioc varietals: Hyperdiversity or social currency.
885 Human Ecology 36:679–697.
- 886 JARVIS, D. I., BROWN, A. H., CUONG, P. H., COLLADO-PANDURO, L., LATOURNERIE-MORENO, L.,
887 GYAWALI, S., TANTO, T., SAWADOGO, M., MAR, I., AND SADIKI, M. 2008. A global perspective of
888 the richness and evenness of traditional crop-variety diversity maintained by farming commu-
889 nities. Proceedings of the National Academy of Sciences 105:5326–5331.
- 890 JOHNS, T., POWELL, B., MAUNDU, P., AND EYZAGUIRRE, P. B. 2013. Agricultural biodiversity as
891 a link between traditional food systems and contemporary development, social integrity and
892 ecological health. Journal of the Science of Food and Agriculture 93:3433–3442.
- 893 JORDANO, P. 1987. Patterns of mutualistic interactions in pollination and seed dispersal: con-
894 nectance, dependence asymmetries, and coevolution. American Naturalist 129:657–677.
- 895 JORDANO, P., BASCOMPTE, J., AND OLESEN, J. M. 2003. Invariant properties in coevolutionary
896 networks of plant-animal interactions. Ecology Letters 6:69–81.
- 897 KARRER, B. AND NEWMAN, M. E. 2011. Stochastic blockmodels and community structure in net-
898 works. Physical Review E 83:016107.
- 899 KERIBIN, C., BRAULT, V., CELEUX, G., AND GOVAERT, G. 2014. Estimation and selection for the
900 latent block model on categorical data. Statistics and Computing pp. 1–16.
- 901 KOLACZYK, E. D. 2009. Statistical analysis of network data: methods and models. Springer Science
902 & Business Media.
- 903 KREMEN, C., WILLIAMS, N. M., AND THORP, R. W. 2002. Crop pollination from native bees at risk
904 from agricultural intensification. Proceedings of the National Academy of Sciences 99:16812–
905 16816.
- 906 LABEYRIE, V., RONO, B., AND LECLERC, C. 2013. How social organization shapes crop diversity:
907 an ecological anthropology approach among Tharaka farmers of Mount Kenya. Agriculture and
908 Human Values 31:97–107.
- 909 LANDE, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple
910 communities. Oikos pp. 5–13.
- 911 LAZEGA, E., MOUNIER, L., SNIJDERS, T., AND TUBARO, P. 2012. Norms, status and the dynamics of
912 advice networks: A case study. Social Networks 34:323–332.

- 913 LECLERC, C. AND COPPENS D'EECKENBRUGGE, G. 2012. Social organization of crop genetic diver-
 914 sity. the $g \times e \times s$ interaction model. Diversity 4:1–32.
- 915 LEGER, J.-B. 2015. blockmodels: Latent and Stochastic Block Model Estimation by a 'V-EM' Algo-
 916 rithm. R package version 1.1.1.
- 917 LEI, J., RINALDO, A., ET AL. 2014. Consistency of spectral clustering in stochastic block models.
 918 The Annals of Statistics 43:215–237.
- 919 MARIAC, C., JEHIN, L., AND SAÏDOU A.-A.A.AND VIGOUROUX, Y. 2011. Genetic basis of pearl millet
 920 adaptation along an environmental gradient investigated by a combination of genome scan and
 921 association mapping. Molecular ecology 20:80–91.
- 922 MEILLEUR, B. A. 1998. Clones within clones: Cosmology and esthetics and Polynesian crop selec-
 923 tion. Anthropologica 40:71–82.
- 924 MIKLÓS, I. AND PODANI, J. 2004. Randomization of presence-absence matrices: comments and
 925 new algorithms. Ecology 85:86–92.
- 926 MULUMBA, J. W., NANKYA, R., ADOKORACH, J., KIWUKA, C., FADDA, C., DE SANTIS, P., AND
 927 JARVIS, D. I. 2012. A risk-minimizing argument for traditional crop varietal diversity use to re-
 928 duce pest and disease damage in agricultural ecosystems of Uganda. Agriculture, Ecosystems &
 929 Environment 157:70–86.
- 930 PADOCH, C. AND JONG, W. D. 1991. The house gardens of Santa Rosa: Diversity and variability in
 931 an Amazonian agricultural system. Economic Botany 45:166–175.
- 932 PERONI, N. AND HANAZAKI, N. 2002. Current and lost diversity of cultivated varieties, espe-
 933 cially cassava, under swidden cultivation systems in the Brazilian Atlantic Forest. Agriculture,
 934 Ecosystems & Environment 92:171–183.
- 935 PERRAULT-ARCHAMBAULT, M. AND COOMES, O. T. 2008. Distribution of agrobiodiversity in home
 936 gardens along the Corrientes River, Peruvian Amazon. Economic Botany 62.
- 937 PODANI, J. AND SCHMERA, D. 2012. A comparative evaluation of pairwise nestedness measures.
 938 Ecography 35:889–900.
- 939 REYES-GARCÍA, V., MOLINA, J. L., CALVET-MIR, L., ACEITUNO-MATA, L., LASTRA, J. J., ONTILLERA,
 940 R., PARADA, M., PARDO-DE SANTAYANA, M., RIGAT, M., AND VALLÈS, J. 2013. 'Tertius gaudens':
 941 germplasm exchange networks and agroecological knowledge among home gardeners in the
 942 Iberian Peninsula. Journal of ethnobiology and ethnomedicine 9:1–11.
- 943 RIVAL, L. AND MCKEY, D. 2008. Domestication and diversity in manioc (*Manihot esculenta* Crantz
 944 *ssp. esculenta*, *Euphorbiaceae*). Current Anthropology 49:1119–1128.
- 945 RODRÍGUEZ-GIRONÉS, M. A. AND SANTAMARÍA, L. 2006. A new algorithm to calculate the nested-
 946 ness temperature of presence–absence matrices. Journal of Biogeography 33:924–935.
- 947 SALICK, J., CELLINESE, N., AND KNAPP, S. 1997. Indigenous diversity of Cassava: Generation,
 948 maintenance, use and loss among the Amuesha, Peruvian upper Amazon. Economic Botany
 949 51:6–19.

- 950 SAMBERG, L., SHENNAN, C., AND ZAVALA, E. 2013. Farmer seed exchange and crop diversity
951 in a changing agricultural landscape in the southern highlands of Ethiopia. Human Ecology
952 41:477–485.
- 953 SUBEDI, A., CHAUDHARY, P., BANAIYA, B. K., RANA, R. B., TIWARI, R. K., RIJAL, D. K., STHAPIT,
954 B. R., AND JARVIS, D. 2003. Who maintains crop genetic diversity and how?: Implications for
955 On-farm Conservation and Utilization. Culture & Agriculture 25:41–50.
- 956 TAPIA, M. E. 2000. Mountain Agrobiodiversity in Peru. Mountain Research and Development
957 20:220–225.
- 958 THÉBAULT, E. AND FONTAINE, C. 2010. Stability of ecological communities and the architecture of
959 mutualistic and trophic networks. Science 329:853–856.
- 960 TUXILL, J., REYES, L. A., MORENO, L. L., UICAB, V. C., AND JARVIS, D. I. 2010. All maize is not
961 equal: maize variety choices and mayan foodways in rural Yucatan, Mexico, pp. 467–486. In Pre-
962 Columbian Foodways: interdisciplinary approaches to food, culture, and markets in Ancient
963 Mesoamerica. Springer, j.e. staller and m.d. carrasco edition.
- 964 ULRICH, W. AND GOTELLI, N. J. 2012. A null model algorithm for presence–absence matrices
965 based on proportional resampling. Ecological Modelling 244:20–27.
- 966 VIGOUROUX, Y., BARNAUD, A., SCARCELLI, N., AND THUILLET, A.-C. 2011. Biodiversity, evolution
967 and adaptation of cultivated crops. Comptes rendus biologies 334:450–457.
- 968 WASSERMAN, S. AND FAUST, K. 1994. Social network analysis: methods and applications. Cam-
969 bridge University Press.
- 970 ZAMAN, A. AND SIMBERLOFF, D. 2002. Random binary matrices in biogeographical ecol-
971 ogy—instituting a good neighbor policy. Environmental and Ecological Statistics 9:405–421.
- 972 ZIMMERER, K. S. 1991. Labor shortages and crop diversity in the Southern Peruvian Sierra.
973 Geographical Review 81:414–432.

974 **Supplementary Material**

975 **Statistical power study of the contribution test**

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

980

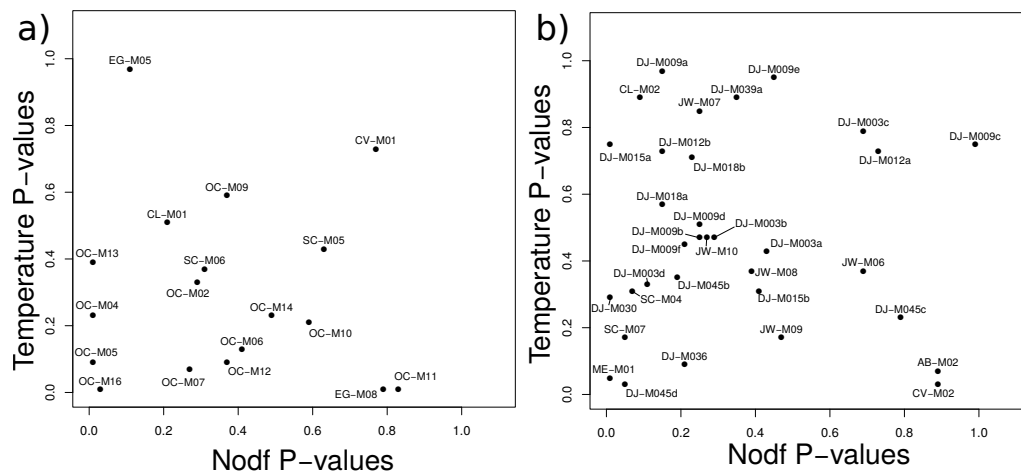


Figure 15: Plot representing in X-axis the NODF P-values computed by re-sampling, and in Y-axis the Temperature P-values computed by re-sampling. In the both cases, re-sampling was performed using the configuration model: a) for data sets collected at the specific level, b) for data sets collected at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

981 Tables P: [ref.](#) present the proportion of rejection (in %) when the α level is set to 1% and
 982 5% for the different sizes of incidence matrices. The tests have globally the same power. When
 983 there are only $n = 20$ households or $m = 20$ plants, the power is quite low. For $n = m = 50$, the
 984 power is totally satisfactory. In the null model when there is no interaction between richness of
 985 the households and the status of plants, the p-values are nearly uniformly distributed on $[0, 1]$ as
 986 expected under a null model.

987 Estimation of nestedness

988 This section describes the nestedness results obtained on the meta-data set using two methods:
 989 the Temperature (Rodríguez-Gironés and Santamaría, 2006) and the NODF (Almeida-Neto et al.,
 990 2008). The figure 15 represents the P-values computed for each estimator after re-sampling using
 991 the configuration model introduce in section 4.4.1. Our results are consistent with Podani and
 992 Schmera (2012) because for the same meta-data set, tests performed with one or another index
 993 were inconsistent.