

# A generative model for reciprocity and community detection in networks

Hadiseh Safdari,<sup>1,\*</sup> Martina Contisciani,<sup>1,†</sup> and Caterina De Bacco<sup>1,‡</sup>

<sup>1</sup>*Max Planck Institute for Intelligent Systems, Cyber Valley, Tuebingen 72076, Germany*

We present a probabilistic generative model and efficient algorithm to model reciprocity in directed networks. Unlike other methods that address this problem such as exponential random graphs, it assigns latent variables as community memberships to nodes and a reciprocity parameter to the whole network rather than fitting order statistics. It formalizes the assumption that a directed interaction is more likely to occur if an individual has already observed an interaction towards her. It provides a natural framework for relaxing the common assumption in network generative models of conditional independence between edges, and it can be used to perform inference tasks such as predicting the existence of an edge given the observation of an edge in the reverse direction. Inference is performed using an efficient expectation-maximization algorithm that exploits the sparsity of the network, leading to an efficient and scalable implementation. We illustrate these findings by analyzing synthetic and real data, including social networks, academic citations and the Erasmus student exchange program. Our method outperforms others in both predicting edges and generating networks that reflect the reciprocity values observed in real data, while at the same time inferring an underlying community structure. We provide an open-source implementation of the code online.

## Introduction

Reciprocity in directed networks, the tendency of a pair of nodes to form mutual connections between each other [1], is an important feature of many social relationships. Its impact ranges from affecting the development of exchange and power to determining the emergence of trust and solidarity [2, 3]. Behavior of this kind has also been found in many kinds of networks that reflect human and institutional interaction, e.g., the world wide web, online dating, interfirm contracts, journal citations and email communication [4–8].

Among the various network modeling approaches, that of probabilistic generative models allows for a rigorous theoretical foundation within the framework of statistical inference, as well as a flexible incorporation of domain knowledge in the modeling assumptions. Here, we consider a latent variable model, a probabilistic approach that contains latent and observed variables. The latent variables encode hidden patterns in the data, such as community memberships, and determine the probability of ties between nodes. For instance, knowing which communities two nodes belong to helps determine the likelihood of their interaction. While in some simple cases, community structure may explain the tendency toward reciprocation [9], this mechanism may not be enough to capture more complex scenarios. Indeed, many generative models with community structure fail to reproduce the values of reciprocity observed in real networks, as we discuss in more details later. Conversely, several models aimed at capturing reciprocity do not account for community structure [10, 11]. It is reasonable to expect that the mechanism regulating the existence of interactions can be influenced by both patterns of communities

and reciprocity. In addition, communities are often interpretable objects and may correspond to functional units, hence the value of including them both in the model formulation. Incorporating these two effects into a coherent latent variable model comes with the main challenge of relaxing the conditional independence assumption between edges, which is commonly made in generative models to ease mathematical derivations. In addition, this task requires properly capturing conditional probabilities, as we describe later. Inspired by these insights, we propose a novel probabilistic latent variable approach to model networks that preserves the benefits of generative models, while capturing both community structure and reciprocity.

Models for reciprocity and latent community structure have largely been developed independently of one another, and only a handful of works have hinted at incorporating them into a unique framework. For instance, Holland et al. [9] describe a stochastic block model with reciprocity, while Garlaschelli and Loffredo [12] point towards a possible relationship between their model for reciprocity and general hidden variable models. Reciprocity is often modeled by means of exponential random graphs [10, 11, 13, 14], where it is treated as a measured network property that needs to be reproduced (often together with other network properties like the degree) by sampling networks using statistical mechanics principles, e.g., maximum entropy. This approach does not include latent variables such as community membership; therefore possible group structures can only be estimated a posteriori on the sampled networks. In contrast, generative models incorporate a priori community structure by means of latent variables, which are inferred from the observed interactions [15, 16]. However, reciprocity is not explicitly included as a mechanism for tie formation, thus these models often fail to reproduce the observed reciprocity values of real networks. Consequently, a generative method whose latent variables describe both reciprocity and community memberships is needed.

\* hadiseh.safdari@tuebingen.mpg.de

† martina.contisciani@tuebingen.mpg.de; Contributed equally.

‡ caterina.debacco@tuebingen.mpg.de

## Results

### Relaxing the conditional independence assumption

A possible explanation for the practical deficiency of generative models with communities to recover observed reciprocity values is the common assumption of conditional independence between edges, which makes the problem both analytically and computationally more tractable. This assumption states that the likelihood of a directed tie between two nodes depends only on their community membership (and other possible model parameters), but not on the existence of a reciprocated edge. This might be too strict of an assumption to capture the feature of reciprocity, where it is reasonable to expect that the existence of an edge in one direction should also be conditioned on the existence of an edge in the opposite direction. For instance, if an author  $i$  has cited another author  $j$ , this would help predict the probability of  $j$  also citing  $i$ . At the same time, knowing the communities that the authors belong to, should also help estimating this probability. Mathematically, this can be translated to relaxing the assumption of conditional independence, which is the approach we take here. Formally, we represent interactions between  $N$  individuals as a weighted asymmetric matrix  $A$ , with entries  $A_{ij}$  being the number (or weight) of interactions from  $i$  to  $j$ ; for instance, the number of favors or services that  $i$  does for  $j$ , or the number of times that  $i$  has endorsed  $j$ , e.g., as paper citations. Our model assigns a *joint* likelihood  $P(A_{ij}, A_{ji}|\Theta)$  to edges involving the same pairs of nodes  $(i, j)$ , given some set of latent parameters  $\Theta$ . This means that we account for interactions between edges connecting the same pair of nodes. Specifically, we assume the likelihood of a network to factorize as:

$$P(A|\Theta) = \prod_{i < j} P(A_{ij}, A_{ji}|\Theta) \quad . \quad (1)$$

This is fundamentally different from the prevalent approaches in generative models, where, typically, one assumes that *individual* edges are conditionally independent given the parameters, i.e.,  $P(A|\Theta) = \prod_{i,j} P(A_{ij}|\Theta)$ . Notice that edges involving different pairs of nodes remain conditionally independent as in standard approaches. Equivalently, in terms of the conditional distribution of an individual edge  $P(A_{ij}|A_{ji}, \Theta)$ , we assume that this can be different than its marginal distribution  $P(A_{ij}|\Theta)$ . To the extent of our knowledge, this assumption has never been deeply questioned, except for a few works [17, 18]. As firstly pointed out by Hoff [19], there are theoretical groundings for this to hold in common scenarios, due to generalizations of de Finetti's theorem by Aldous [20] and Hoover [21] (see [18] for a detailed discussion). They show that, for exchangeable graphs, i.e., networks where there is no natural ordering between nodes (which is often the case), the joint probability function of the adjacency entries can be properly represented using

latent variables on nodes and pairs, i.e., the joint can be factorized as a product on edges given the latent variables. However, in the case of directed networks, where the adjacency matrix is asymmetric, as in our case, a precise representation can only be obtained using Eq. (1). While standard conditionally independent models can in principle arbitrarily well approximate the whole network distribution [22], in practice it is not known how state-of-the-art models perform on this regard. To effectively model reciprocity, we relax the assumption of conditional independence and include the pairwise dependencies of two directed edges between pairs of nodes; such minimal relaxation is required to effectively model reciprocity. We compare results against standard conditionally independent models in terms of various performance metrics on both synthetic and real data.

### The community-reciprocity model

To fully specify the joint likelihood in Eq. (1) we need to characterize conditional distributions and one-point marginals like  $P(A_{ij}|A_{ji}, \Theta)$  and  $P(A_{ij}|\Theta)$ . This can be done by formalizing the intuition of what drives the formation of edges based on the application at hand. Here, we aim at capturing reciprocity, hence we assume that observed interactions exist because of two types of contributions: the communities that nodes belong to, as in general community detection frameworks like the stochastic block model [9], and the fact that an individual that receives a directed interaction is more likely to reciprocate. In order to construct a model flexible enough to capture weighted networks and overlapping communities, we utilize a mixed-membership approach, similar to [15, 16], to model how communities regulate edge formation.

Given the adjacency matrix  $A$ , our goal is to find community memberships of nodes and the magnitude of the reciprocity effect in the network. Bringing the contributions of reciprocity and community structure together, we model the conditional probability of  $A_{ij}$  given  $A_{ji}$  as drawn from a Poisson distribution

$$P(A_{ij}|A_{ji}, \Theta) = \frac{e^{-\lambda_{ij}} \lambda_{ij}^{A_{ij}}}{A_{ij}!} \quad , \quad (2)$$

with mean

$$\lambda_{ij} = \lambda_{ij}^0 + \eta A_{ji} = \left( \sum_{k,q=1}^K u_{ik} v_{jq} w_{kq} \right) + \eta A_{ji} \quad , \quad (3)$$

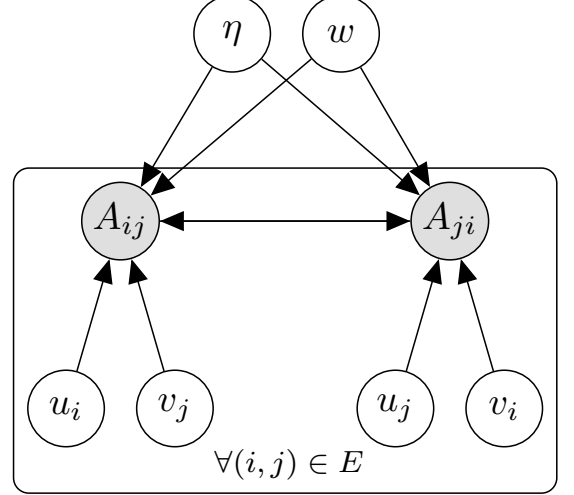
where  $u_{ik}$ ,  $v_{ik}$  are entries of  $K$ -dimensional vectors  $u_i$  and  $v_i$ , the out-going and in-coming communities respectively;  $w_{kq}$  are the entries of a  $K \times K$  affinity matrix, which regulates the structure of communities, e.g., assortative when its diagonal entries are greater than off-diagonal entries, in this case edges are more likely between nodes in the same community;  $\eta$  is the reciprocity parameter, and it regulates the impact of observing  $A_{ji}$

to predict the existence of  $A_{ij}$ . Thus  $\lambda_{ij}$  includes separate contributions from both community parameters and reciprocity coefficient. Intuitively, an edge with weight  $A_{ij}$  exists if  $i$  and  $j$  belong to compatible communities (compatibility is regulated by the affinity matrix) or because of the reciprocity effect of observing the opposite edge  $A_{ji}$ . For instance, an author might cite another one because they belong to the same community (e.g., a research sub-field) or because she was cited by the other on a previous publication. We denote with  $\Theta = (u, v, w, \eta)$  the latent parameters that we want to infer. We omit from it the number of communities  $K$ , as in this work we assume this as given. When unknown, as in our experiments with real data, we estimate it by using cross-validation. Notice that this formulation assumes additive contributions, which implies that we can have zero contribution from one term and still observe the existence of an edge because of the other term. If both are non-zero, their total impact sums up. This is conceptually different than a multiplicative contribution, a possible modeling choice that we do not explore here. Finally, as we need positive  $\lambda_{ij}$ , we assume  $\eta \geq 0$ . This restricts the model to have positive reciprocity contribution, i.e., receiving an in-coming edge can only boost the likelihood of the corresponding out-going edge, but not decrease it. Although this assumption could be limiting in certain contexts, it nevertheless applies to several relevant scenarios, in particular to the cases we study here. Relaxing this assumption, and suitably modifying the underlying theoretical model is left for future work. Our model specifies conditional probabilities, however, we do not assume the existence of a consistent joint distribution. In fact, finding a closed-form for the joint in Eq. (1) consistent with our proposed conditional requires specifying a marginal probability function and then enforce consistency equations like  $\sum_{A_{ji}} P(A_{ij}|A_{ji}, \Theta) P(A_{ji}|\Theta) = P(A_{ij}|\Theta)$ . Depending on the choice of this marginal, enforcing consistency might be non-trivial, as it may require performing intractable marginalization. Early formalizations of the consistency between conditional and joint distribution has been provided by Besag's Auto-Poisson models [23]. In the context of graphical models, few models specify conditional Poisson distributions [24, 25], but without considering latent variables. In the absence of a closed-form joint distribution, we adopt a tractable pseudo-likelihood approach [23], where instead of optimizing the exact likelihood of Eq. (1), we consider the approximation:

$$P(A|\Theta) = \prod_{i < j} P(A_{ij}, A_{ji}|\Theta) \approx \prod_{i, j} P(A_{ij}|A_{ji}, \Theta) \quad , \quad (4)$$

which is available in closed-form as it requires only the conditional probabilities, which we specified above. The equality holds only when  $A_{ij}$  and  $A_{ji}$  are conditionally independent, the common assumption in network generative models, as in that case  $P(A_{ij}|A_{ji}, \Theta) = P(A_{ij}|\Theta)$ . This is not our case since we relax this assumption and

Eq. (4) is only an approximation. This approach has also been considered in dyadic-dependent models [26], for community detection in networks [27], and for local Poisson graphical models [24]. A visual overview of our model is shown in Fig. 1.



**FIG. 1: Graphical model representation.**  $A_{ij}$  and  $A_{ji}$  are the edges involving the same pairs of nodes  $(i, j)$ ;  $\eta, w, u, v$  are the latent parameters  $\Theta$ ;  $E$  denotes the set of network edges.

### Inference with expectation-maximization

The goal is to find the community and reciprocity parameters, i.e.,  $\Theta$ , given the adjacency matrix. Defining  $L_{ij}^{ps}(\Theta, A_{ji}) = \log P(A_{ij}|A_{ji}, \Theta)$  and neglecting the factorial term which is independent of these parameters, we have the log-pseudo-likelihood:

$$L^{ps}(\Theta) = \sum_{i, j} L_{ij}^{ps}(\Theta) = \sum_{i, j} (A_{ij} \log \lambda_{ij} - \lambda_{ij}) \quad . \quad (5)$$

We aim at maximizing this quantity, but the presence of the logarithmic term makes this maximization difficult. However, using a variational approach by means of Jensen's inequality, it can be shown (see Supplementary Information Sec. S1 A) that maximizing  $L^{ps}(\Theta)$  is

equivalent to maximizing

$$L^{ps}(\Theta, \rho, \phi) = \sum_{i,j} \left\{ A_{ij} \rho_{ij}^{(1)} \left( \sum_{k,q} \phi_{ijkq} \log u_{ik} v_{jq} w_{kq} - \sum_{k,q} \phi_{ijkq} \log \phi_{ijkq} \right) + A_{ij} \rho_{ij}^{(2)} \log \eta A_{ji} - A_{ij} \left( \rho_{ij}^{(1)} \log \rho_{ij}^{(1)} + \rho_{ij}^{(2)} \log \rho_{ij}^{(2)} \right) - \sum_{k,q} u_{ik} v_{jq} w_{kq} - \eta A_{ji} \right\}, \quad (6)$$

with respect to  $\Theta$ ,  $\rho = (\rho^{(1)}, \rho^{(2)})$ , and  $\phi$ , where

$$\rho_{ij}^{(1)} = \frac{\lambda_{ij}^0}{\lambda_{ij}^0 + \eta A_{ji}}, \quad \rho_{ij}^{(2)} = \frac{\eta A_{ji}}{\lambda_{ij}^0 + \eta A_{ji}}, \quad (7)$$

$$\phi_{ijkq} = \frac{u_{ik} v_{jq} w_{kq}}{\lambda_{ij}^0}. \quad (8)$$

Constraints to the parameters can be arbitrarily added, e.g.,  $\sum_k u_{ik} = \sum_k v_{ik} = 1$ , by incorporating Lagrange multipliers inside Eq. (5) and repeating similar calculations, see Supplementary Information Sec. S1B for more details. In our numerical experiments, we consider both constrained and unconstrained cases. We can perform this optimization by alternatively updating the various parameters, with an expectation-maximization (EM) algorithm. At each step, one updates  $\rho$  and  $\phi$  using Eqs. (7)-(8) (E-step) and then maximizes  $L^{ps}(\Theta, \rho, \phi)$  with respect to  $\Theta$  by setting partial derivatives to zero (M-step). For instance, the update equation for  $\eta$  is obtained by considering the partial derivative:

$$\frac{\partial L^{ps}}{\partial \eta} = \sum_{i,j} \left[ \frac{A_{ij} \rho_{ij}^{(2)}}{\eta} - A_{ji} \right]. \quad (9)$$

Setting this to zero and defining  $M = \sum_{i,j} A_{ij}$ , we obtain:

$$\eta^{(t+1)} = \frac{\sum_{i,j} A_{ij} \rho_{ij}^{(2)}}{\sum_{i,j} A_{ij}} = \frac{\eta^{(t)}}{M} \sum_{i,j} \frac{A_{ij} A_{ji}}{\lambda_{ij}^{(t)}}. \quad (10)$$

The whole routine is described in Algorithm 1 and the detailed derivations are in the Supplementary Information Sec. S1. This algorithm is computationally efficient and scalable to large system sizes as it exploits the sparsity of the dataset. Indeed, all the updates involve in the numerator sums over  $A_{ij}$ , hence only the non-zero entries count, giving an algorithmic complexity of  $O(M K^2)$ .

---

**Algorithm 1** CRep: EM algorithm

---

**Input:** network  $A = \{A_{ij}\}_{i,j=1}^N$ ,  
number of communities  $K$

**Output:** membership vectors  $u = [u_{ik}]$ ,  $v = [v_{ik}]$ ;  
network-affinity matrix  $w = [w_{kq}]$ ; reciprocity  
parameter  $\eta$ .

Initialize  $u, v, w, \eta$  at random.

Repeat until convergence:

1. Calculate  $\rho^{(1)}$  and  $\phi$  (E-step):

$$\rho_{ij}^{(1)} = \frac{\lambda_{ij}^0}{\lambda_{ij}^0 + \eta A_{ji}}, \quad \phi_{ijkq} = \frac{u_{ik} v_{jq} w_{kq}}{\lambda_{ij}^0}$$

2. Update parameters  $\Theta$  (M-step):

i) for each node  $i$  and community  $k$  update memberships:

$$u_{ik} = \frac{1}{\gamma_i^u} \sum_{j,q} A_{ij} \rho_{ij}^{(1)} \phi_{ijkq}$$

$$v_{ik} = \frac{1}{\gamma_i^v} \sum_{j,q} A_{ji} \rho_{ji}^{(1)} \phi_{jiqk}$$

ii) for each pair  $(k, q)$  update affinity matrix:

$$w_{kq} = \frac{\sum_{i,j} A_{ij} \rho_{ij}^{(1)} \phi_{ijkq}}{\sum_{i,j} u_{ik} v_{jq}}$$

iii) update reciprocity parameter:

$$\eta = \frac{\eta^{(t)}}{M} \sum_{i,j} \frac{A_{ij} A_{ji}}{\lambda_{ij}^{(t)}}$$

Note:  $\gamma_i^u, \gamma_i^v$  are quantities that are defined differently for constrained and unconstrained values of  $u_i$  and  $v_i$ . In the constrained case, they correspond to Lagrange multipliers; see Supplementary Information Sec. S1B.

---

### A benchmark generative model with communities and reciprocity

So far we have focused on recovering the parameters given the data, i.e., the inference. In this section, instead, we propose a benchmark probabilistic generative model to generate synthetic data with intrinsic community structure and a given reciprocity value. It takes as input a set of membership vectors,  $u_i$  and  $v_i$ , affinity matrix  $w$  and reciprocity parameter  $\eta$ ; the output is a directed network with adjacency matrix  $A$ . In this formulation, edges between a given pair of nodes are generated stochastically; one edge being generated first and independent from the other, while the formation of the opposite edge depends on how the first was drawn. The pairs of edges are conditionally independent from each other. Formally, we aim at sampling pairs of edges from Eq. (1), which can be done by selecting a marginal  $P(A_{ij}|\Theta)$  and a conditional distribution  $P(A_{ji}|A_{ij}, \Theta)$ . By assuming a Poisson conditional as in Eq. (2) and a



Poisson marginal, our model would reduce to a standard (conditionally independent) generative model with communities in the case of zero reciprocity parameter. Even though with this choice the joint is not tractable, this is not an issue, as we do not aim to use the joint to compute quantities analytically, but rather focus on sampling from it. Formally, given the input set of latent variables  $\Theta = (u, v, w, \eta)$ , we draw a pair  $(A_{ij}, A_{ji})$  consistently with the joint  $P(A_{ij}, A_{ji}|\Theta)$ , in a two-step sampling routine:

1. Select with a coin-flip one direction,  $(i, j)$  or  $(j, i)$ . Say we select  $(i, j)$ .
2. Sample  $A_{ij}$  from the marginal

$$P(A_{ij}|\Theta) = \text{Pois}(A_{ij}; m_{ij}) \quad , \quad (11)$$

where

$$m_{ij} = \frac{\lambda_{ij}^0 + \eta \lambda_{ji}^0}{(1 - \eta^2)} \quad (12)$$

is the mean of the joint distribution consistent with the conditional probability (see Supplementary Information Sec. S1 C).

3. Sample  $A_{ji}$  from the conditional

$$P(A_{ji}|A_{ij}, \Theta) = \text{Pois}(\lambda_{ji}^0 + \eta A_{ij}) \quad , \quad (13)$$

using the previously extracted value of  $A_{ij}$ .

The Poisson distribution may generate multiple edges between a pair of nodes, so this model may create multi-graphs. This is consistent with the interpretation that  $A_{ij}$  is the number, or total weight, of links from  $i$  to  $j$ . If we wish to generate binary networks where  $A_{ij} \in \{0, 1\}$ , we use the fact that the Poisson and Bernoulli distributions become close in the sparse limit. To enforce sparsity, it is sufficient to multiply the  $\lambda_{ij}^0$  by a constant  $c$ , as the  $m_{ij}$  in Eq. (12) will also be automatically rescaled by the same quantity. The constant can be fixed by choosing a value for the expected number of (weighted) edges:

$$\mathbb{E}[M] = \sum_{i,j} \frac{c \lambda_{ij}^0 + c \eta \lambda_{ji}^0}{1 - \eta^2} = \frac{c}{1 - \eta} \sum_{i,j} \lambda_{ij}^0 \quad (14)$$

$$\rightarrow c = (1 - \eta) \frac{\mathbb{E}[M]}{\sum_{i,j} \lambda_{ij}^0} \quad . \quad (15)$$

Imagine now a practitioner willing to control for the relative contribution of community and reciprocity in generating edges. Our model naturally allows this possibility, as this is encoded by  $\eta$ . To see this explicitly, we calculate the fraction of edges generated by community effects only:

$$cr_{ratio} := \frac{\sum_{i,j} \lambda_{ij}^0}{\mathbb{E}[M]} = 1 - \eta \quad , \quad (16)$$

where we used Eq. (12) to rewrite the denominator. Thus, by varying  $\eta$  in the input, one automatically tunes the interplay community *vs* reciprocity:  $\eta$  close to 0 gives a network whose edges depend mostly on the community structure imposed by the membership vectors; instead,  $\eta$  close to 1 results in a network with lower impact of community structure, i.e., reciprocity has also significant impact on edge formation. Notice that it is not possible to have a contribution purely due to reciprocity, as this phenomenon implicitly requires the existence of another mechanism to produce one of the two possible edges, here the community structure. This can also be seen by observing that Eq. (12) can be rewritten as  $m_{ij} = \lambda_{ij}^0 + \frac{\eta}{1-\eta^2} (\eta \lambda_{ij}^0 + \lambda_{ji}^0)$ ; while the first term only depends on communities, the second term depends on both communities and reciprocity and they cannot be separated independently.

### Predicting network reciprocity

In directed networks, reciprocity  $r$  is usually defined as the fraction of edges that are reciprocated [1], although other definitions exist to capture this feature [14, 28]. With our probabilistic model, we can compute the expected value of a related quantity

$$r_w := \frac{\sum_{i,j} [A_{ij} A_{ji}]}{\sum_{i,j} [A_{ij}]} \quad , \quad (17)$$

which corresponds to reciprocity in the case of binary adjacency matrices. A natural question is thus how this observable quantity is related to the reciprocity parameter  $\eta$ . In fact, we show that, provided some assumptions for the second moment  $\mathbb{E}[A_{ij}^2]$  (see Supplementary Information Sec. S1 D),  $\eta$  is a lower bound for it:

$$\mathbb{E}[r_w] = \eta + \frac{\sum_{i,j} [\lambda_{ij}^0 m_{ji} + \eta m_{ji}^2]}{\sum_{i,j} m_{ij}} \geq \eta \quad . \quad (18)$$

The tightness of this bound depends on the latent variables through  $\lambda_{ij}^0$ , (implicitly)  $m_{ij}$ , and  $m_{ji}$ . Empirically, we find that in the majority of the experiments the bound is very tight, i.e.,  $\mathbb{E}[r_w] \approx \eta$  and the other terms in Eq. (18) are much smaller than  $\eta$  in models with the conditional independence assumption, such as our proposed model with  $\eta = 0$ ,  $\mathbb{E}[r_w] = \frac{\sum_{i,j} m_{ij} m_{ji}}{\sum_{i,j} m_{ij}}$ . In fact, in these models, the term  $\sum_{i,j} m_{ij} m_{ji}$  is often very small – we show empirical evidence of this later; therefore, even in networks with high reciprocity, they could poorly recover it. The empirical indication for practitioners is that networks generated by models with the conditional independence assumption have reciprocity values significantly different from those observed in real data.

## Predicting reciprocated edges

The dependence structure between pairs of edges should allow us to predict the existence of a reciprocated tie *if* an edge in the opposite direction is observed, such as the citation of a paper if an author has been cited before by someone else. This is a kind of link prediction task, which lets us test the dependence assumption. It is also a principled way of comparing the accuracy of various generative models for datasets where no ground truth for the latent variables is known.

Conditional edge prediction can be formulated as follows: what is the probability of an edge  $i \rightarrow j$  conditioned on observing the opposite existing edge (or non-existing edge)  $j \rightarrow i$ ? Our model naturally outputs this conditional probability. In contrast, a generative model that assumes conditional independence between edges is not capable of exploiting this extra information. It could only estimate marginal probabilities that do not depend on observing the opposite edge as it uses only the parameters such as community memberships and the affinity matrix. Our model is not capable of fully estimating marginal distributions but nevertheless can estimate its expected value as in Eq. (12). This is often the main quantity used in prediction tasks, as it plays the role of a score for estimating the entries  $A_{ij}$ . Therefore, with our model we can also predict regular edge existence, where we simply aim at predicting an edge without any extra information but the inferred parameters.

In our experiments below, we test various generative models for both regular and conditional edge prediction by giving them access to 80% of the possible pairs of nodes in the network (the training data) and then asking them to predict the existence of one edge in the pair, given observing the value of the other of the remaining pairs of nodes (the test data). As performance metrics, we measure the AUC on the test data, i.e., the probability that a randomly selected edge has higher expected value than a randomly selected non-existing edge. We compute both the regular AUC by using as a score the expected value  $\mathbb{E}_{P(A_{ij}|\Theta)}[A_{ij}]$  and the *conditional* AUC (AUC-cond), which uses  $\mathbb{E}_{P(A_{ij}|A_{ji},\Theta)}[A_{ij}]$  as the score, i.e., the expected value over the conditional distribution. The latter can only be computed for our algorithm, as for the others the marginal distribution is the same as the conditional, and thus the two AUC values coincide, see Sec. “Materials and methods” for more details.

## Results on real and synthetic data

We now demonstrate our model by applying it to both real and synthetic data. In the real-world datasets available to us, we only have a directed network of observed interactions, i.e., there is no available ground truth for the actual membership and reciprocity parameters; consequently, their relative contributions in edge formation cannot be tuned. Thus, we first validate our

model and competing algorithms on synthetic data produced with different generative models, testing the ability to: i) generate sample networks that replicate relevant network quantities like reciprocity, similar to those observed on the input networks; ii) perform edge prediction tasks. We then investigate our model’s performance on real-world datasets. In the tests below, we use our model in various ways: the constrained version with constraints on the membership parameters  $u$  and  $v$  such that  $\sum_k u_{ik} = \sum_k v_{ik} = 1, \forall i$  (CRep), the non constrained version (CRep<sub>nc</sub>), and our model with  $\eta = 0$  (CRep<sub>0</sub>), i.e., without considering the reciprocity effect. For comparison, we use two generative models with latent variables: a community detection-only generative model with a Maximum Likelihood approach [15] (MT), which was the inspiration for the building block of our model in the case  $\eta = 0$ , and a Bayesian Poisson matrix factorization (BPMF) commonly used in recommendation systems [29]. For the edge prediction task on real data, we also consider a supervised learning link-prediction routine (OLP) with topological predictors and the implementation of Ghasemian et al. [30] (see Supplementary Information Sec. S4 C for details).

## Performance for synthetic networks

We study various types of synthetic networks, generated by three different models to cover several network topologies. Two of them cover the extreme scenarios of networks generated, accounting only for community structure or only for reciprocity. For the former we use the standard stochastic block model (SBM) [9] and for the latter the reciprocity model of Holland and Leinhardt (HL) [10]. Our model, instead, is designed to tune the relative impact of community structure and reciprocity in determining edges, by varying the parameter  $\eta$ . Thus, we use the benchmark generative model described above to interpolate between these two extremes by tuning  $\eta$ : for small values we recover the stochastic block model, whereas for higher values we recover a structure similar to Holland and Leinhardt’s model. The generative process is described in detail in the Sec. “Materials and methods”. As a remark, the exact joint likelihood of CRep is not determined in closed-form, however all the models used here for comparison adopt either its Poisson conditional distribution (our model with  $\eta > 0$ ) or its Poisson marginal distribution (all the other models). Thus experiments here are aimed at highlighting differences in the various models’ assumptions. By varying the network sparsity and the impact of communities and reciprocity, we illustrate types of structure that may exist in real-world data, and test each algorithm’s robustness against them on various tasks including edge prediction and the ability to reproduce sample networks that replicate relevant network quantities.

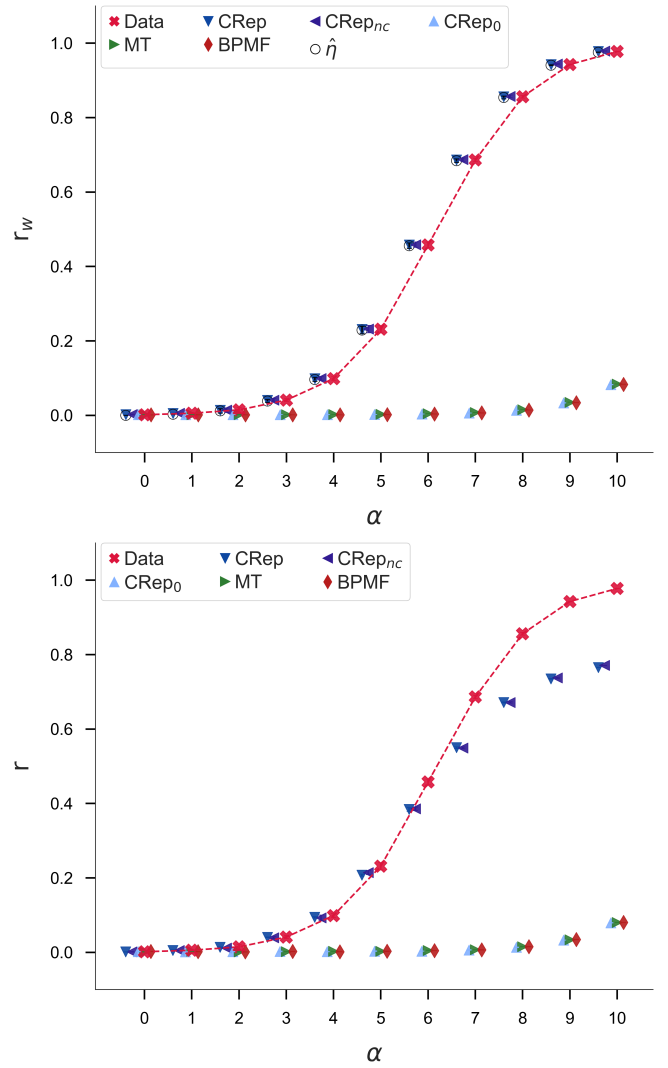
*Recovering topological properties* An important property of a model is the ability to generate network samples

that resemble what is observed in real data. We test this ability by considering topological properties like degree distribution, reciprocity, and hierarchical structure; we calculate their values on network samples which are generated with the various generative models by using the inferred parameters from the given input data. Specifically, we consider networks generated synthetically as explained above, and for each individual network we infer the parameters by each model, and use them to generate five network samples. We compare topological properties of these samples with those observed on the ground truth networks used to infer the parameters.

In particular, we are interested in measuring reciprocity, as the networks generated by algorithms only based on community structure are not capable of reflecting the observed value of the reciprocity in the ground truth network, a shortcoming of these models which indeed limits their applications. The empirical evidence of this observation was part of the motivation to study this problem. In the experiments, we use the standard definition of reciprocity  $r$ , i.e., the ratio of the number of edges pointing in both directions to the total number of edges in the graph (we use the `python` implementation in `networkx`). As anticipated, in networks generated with the stochastic block model,  $r$  is often close to 0, see Supplementary Information Sec. S3 A. Instead, a more interesting scenario is that of networks generated with the main purpose of recovering reciprocity, as in the HL model. This is an example of an exponential random graph model where reciprocity and sparsity are the two topological properties controlled in input. It is also one of the few cases where this type of models is analytical, see Supplementary Information Sec. S2. In this model,  $r$  is tuned by a parameter  $\alpha$  so that the higher its value, the higher the reciprocity. Notice that, as usual in exponential random graphs models, latent variables such as communities are not considered. This model generates unweighted networks, hence  $r \equiv r_w$ . Fig. 2 shows that CRep significantly outperforms all the other generative models in recovering  $r_w$  (top) and  $r$  (bottom), as measured on the sampled networks. The gap between the values of  $r$  and  $r_w$  on the sampled networks is due to the mismatch between the binary adjacency matrices of the networks generated with the HL model (input data) and the weighted sampled ones generated with the various generative models, which use Poisson distributions.

Similar results are obtained for the networks generated with our benchmark generative model. Also in this case, CRep captures reciprocity significantly better than the other models, consistently across values of input  $\eta$  but also for fixed  $\eta$  but different sparsity and degree of overlapping communities. We leave details in the Supplementary Information Sec. S3 A.

We turn now our attention to topological properties other than reciprocity, to investigate how these generative models perform in recovering various relevant properties that might be of interest for a practitioner. Indeed, other possible mechanisms underlying network interac-



**FIG. 2: Reciprocity in HL networks.** Synthetic networks with  $N = 1000$  nodes generated with the model proposed by Holland and Leinhardt by varying the reciprocity parameter  $\alpha$ . Results are empirical averages and standard deviations over 15 samples of three independent synthetic networks (5 samples per input network). The red markers indicate the average on the three input networks. (top) The quantity  $r_w$  as defined in Eq. (17); the empirical average over the samples and the theoretical expectation as in Eq. (18) coincide, hence we omit the markers for the empirical value;  $\hat{\eta}$  is the inferred parameter in CRep and CRep<sub>nc</sub>. (bottom) Standard reciprocity  $r$ . Notice that  $r \equiv r_w$  for the input data, but this is not true for the samples, as the generative models considered here generate weighted edges, i.e. the matrix  $A$  is in general not binary. Error bars are smaller than marker size. Unless otherwise stated, this will be the case in all of the figures.

tions are those that involve more than two individuals (which is the case for reciprocity), namely clustering and hierarchical structure. The first describes the tendency of nodes to form edges within the same neighborhood, hence it involves at least three people. The second involves the notion that the directionality of an unreciprocated edge indicates a hierarchical difference, and this is a global measure in that it requires the whole network for its computation. In our experiments, we find that all models are able to retrieve the degree distribution with good accuracy, while we cannot draw any conclusion about the recovered clustering coefficients, as we observe that the empirical values are too small to notice significant differences (we do not report them here). We focus instead on recovering ranking of nodes, an application relevant when nodes have a score representing some intrinsic notion of relative strength or prestige. For this, we use SpringRank [31], an algorithm for inferring hierarchies in directed networks that assigns real-valued scores to nodes; we calculate the Gini index on these scores to provide a global measure for the whole network; a high Gini index means strong hierarchy, while a low value means that the network is less hierarchical. Comparing the average over the five samples, we find that CRep and CRep<sub>0</sub> are able to perfectly retrieve the Gini index of the original network, while the other models tend to overestimate it, see Supplementary Information Sec. S3 A. This is consistent over the various synthetic network topologies. Notice that this topological property is not influenced by the value of  $\eta$  or the fraction of nodes with mixed-membership used to generate networks, while it decreases as the average degree and  $\alpha$  increase.

*Edge prediction* We test the algorithms’ ability in edge prediction tasks, specifically conditional and regular edge prediction. As we can see from Fig. 3, our model outperforms the others in conditional edge prediction, showing that it is able to efficiently exploit the additional information about the existence of the opposite edge. The performance gap increases with  $\eta$ , as for high values the reciprocity plays a bigger role in edge formation. In the opposite scenario of low  $\eta$ , the impact of reciprocity becomes negligible compared to community structure, and in this case we recover the same results as for the other algorithms. This is expected as our model infers small values of  $\eta$  in this case, thus in practice reducing to a conditional independent model as the others. Performance in terms of regular edge prediction is comparable to the other algorithms for small  $\eta$ , while it drops for intermediate values and then increases again as  $\eta$  grows. CRep and CRep<sub>nc</sub> are very robust in terms of conditional edge predictions as we vary the fraction of nodes with overlapping community membership and the average degree, while fixing  $\eta = 0.5$ . Indeed, their performances are significantly better than that of the other algorithms and do not decrease with increasing overlapping communities and sparsity. Moreover, we find more stable results also in terms of regular edge prediction, where CRep and CRep<sub>nc</sub> have constant values across the

different input parameters, outperforming other methods in critical ranges, e.g., small average degree or high overlap between communities. We leave details in Supplementary Information Sec. S3 B. These synthetic tests suggest that working with conditional probabilities results in more robust estimates of the probability that an edge exists if we have access to the edge in the opposite direction. Performance improvement is more significant when community structure is not the predominant mechanism in edge formation.

For sake of completeness, we also validate the model on community detection tasks and observe good performance of CRep in recovering communities when reciprocity has intermediate or low level. Notice, however, that community detection is not the main focus of our model, as we expect the impact of communities in determining the likelihood of an edge to decrease as reciprocity increases, see Supplementary Information Sec. S3 C. In fact, lower performance in community detection might signal the presence of other mechanisms for edge formation, like that played by reciprocity. Hence, perfect recovering of communities should not be expected in these cases.

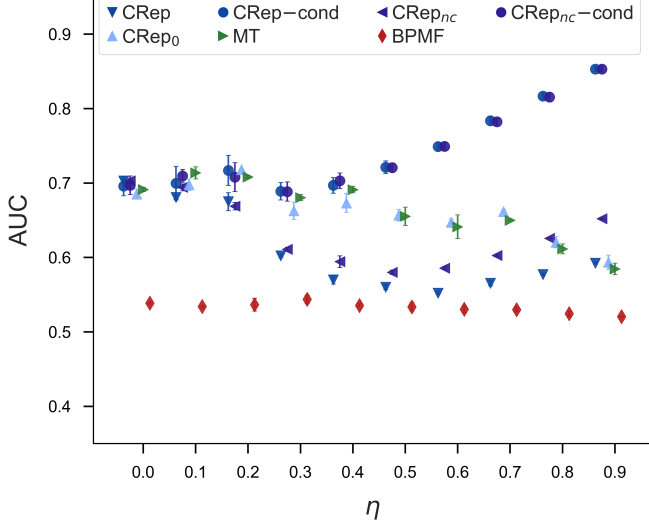
To summarize results on synthetic networks, CRep is capable of suitably capturing the reciprocity values observed in a given network, while also retrieving hierarchical structures. Furthermore, CRep exploits the availability of extra information in performing edge prediction, by increased performance and robustness across various parameters’ ranges.

### Performance for real networks

Above, we evaluated the ability of our model, CRep, to generate network samples that have reciprocity values as expected in input and tested its performance in edge prediction. In this section, we examine these abilities on real world datasets. We apply our method to datasets from a diverse set of fields, with sizes ranging up to  $N \sim 10^4$  nodes and up to  $E \sim 10^5$  links (see Table S1 and Supplementary Information Sec. S4 A for details). Together, these examples cover various types of social relationships, communication interactions, transportation systems, and patterns of citations.

*Recovering topological properties* We apply the same procedure as before to infer the parameters  $\Theta = (u, v, w, \eta)$  from data (this time, real networks) and then generate synthetic network samples based on them. Also in this case, CRep greatly outperforms the other models in recovering  $r$ , consistently across datasets. We show as an example in Fig. 4 the results on the Erasmus dataset (Erasmus Mobility Network 2014 – 2018) [32], and we leave the others in the Supplementary Information Sec. S4 B.

We cannot draw general conclusions in terms of recovering the clustering coefficient and the Gini index of SpringRank scores, as results vary widely depending on

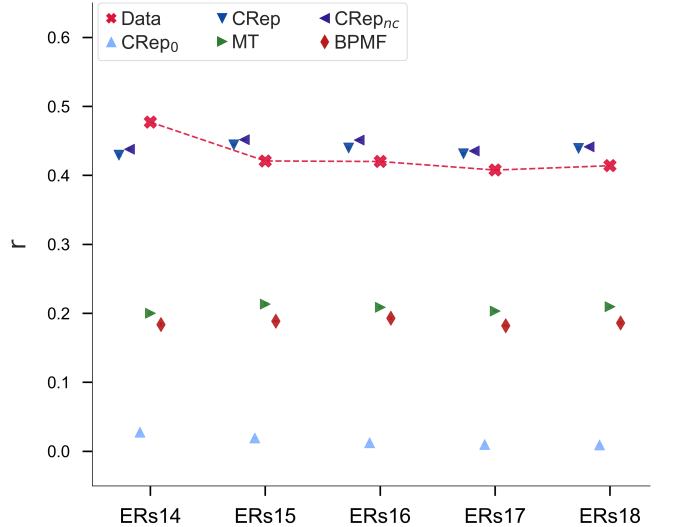


**FIG. 3: Edge prediction in benchmark networks.** Synthetic networks with  $N = 2100$  nodes and  $K = 3$  communities of equal-size unmixed group membership generated with the benchmark generative model proposed above by varying the reciprocity parameter  $\eta$ . The results are averages and standard deviations over three independent synthetic networks and over 5-folds of cross-validation test sets. The accuracy of edge prediction is measured with AUC and the baseline is the random value 0.5.

the datasets, see Supplementary Information Sec. S4B and Fig. S6. As we mentioned before, these are topological properties that involve more complex interactions than pairwise, as in the case of reciprocity (clustering involves triangles and SpringRank score is a global measure). This suggests that, in order to have better performance, one would need to develop more complex models, for instance extending the ideas behind CRep to capture triadic interactions, possibly guided by domain-knowledge about how triadic interactions and reciprocity are related [33]. We leave this for future work, noting that while exponential random graph models can do this, they do not include latent community structure (analogously as for reciprocity).

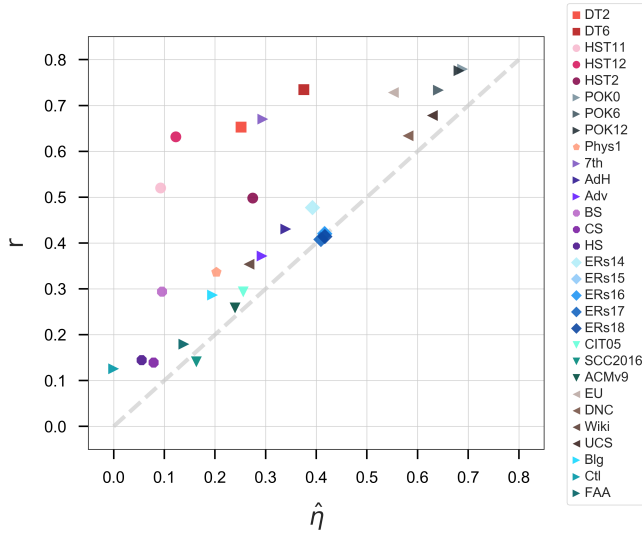
Previously, we have discussed network-related quantities controlled by  $\eta$ , such as the expected fraction of edges purely due to communities ( $cr_{ratio}$ ) or the quantity  $r_w$ . Here we illustrate how the various real networks differ in the inferred values of  $\eta$ , which we denote as  $\hat{\eta}$ . In particular, we show in Fig. 5 how  $\hat{\eta}$  varies according to the reciprocity of these networks, unveiling a non-trivial pattern. While we see a general trend of  $\hat{\eta}$  increasing with  $r$ , there are interval ranges of  $r$  for which  $\hat{\eta}$  varies widely across networks, and vice-versa. For example, we see that for  $r \in [0.6, 0.8]$ ,  $\hat{\eta}$  ranges in  $[0.1, 0.7]$ . This high variability suggests that  $r$  is the result of a complex combination of communities and reciprocity. We notice, for instance, that for high school friendship networks (HST and DT),  $\hat{\eta}$  is low (i.e., in  $[0.1, 0.3]$ ), showing that many reciprocated

edges are explained by community structure. Instead, for online dating (POK) and communication networks (EU and DNC), we observe high values of  $\hat{\eta}$ , signaling a lower impact of communities, as reciprocity plays a bigger role. This reinforces the need to include in network models both mechanisms for explaining edge formation. Notice that these results are possible not only because our model accounts for reciprocity through an explicit parameter  $\eta$ , but also because it infers reciprocity values close to the observed ones, while the other methods fail at this, see Fig. S5.



**FIG. 4: Reciprocity in the Erasmus datasets.** Results are averages and standard deviations of  $r$  over 5 samples generated with the various generative models. The algorithms use the inferred  $\eta$  and community parameters of the dataset – Erasmus in this plot – to generate synthetic network samples. Red markers indicate the values of  $r$  in the real datasets.

*Edge prediction* In the absence of ground truth, as in most real world networks, we test the ability in edge prediction by cross-validation, as done for synthetic networks. Table S3 shows the results in terms of AUC for the generative models CRep, MT, BPF, as well as for OLP; the latter is a type of supervised learning technique which uses network topological information as features to predict the entries of  $A$ . CRep and OLP show the best results, with CRep having high performance for social networks. However, if we consider the conditional AUC, then CRep outperforms all the others in the majority of the datasets, as also observed in synthetic data. This confirms the ability of our model to efficiently exploit the additional information from the adjacency matrix to boost performance in terms of edge prediction.



**FIG. 5: Reciprocity and  $\hat{\eta}$ .** Scatterplot with observed reciprocity (y-axis) and  $\hat{\eta}$  inferred in CRep (x-axis); points are individual real datasets. The dashed grey line indicates the perfect correspondence between  $\eta$  and  $\hat{\eta}$ . Marker shape denotes the type of network as defined in Table S1.

#### Case study: application of CRep to the Erasmus student exchange network

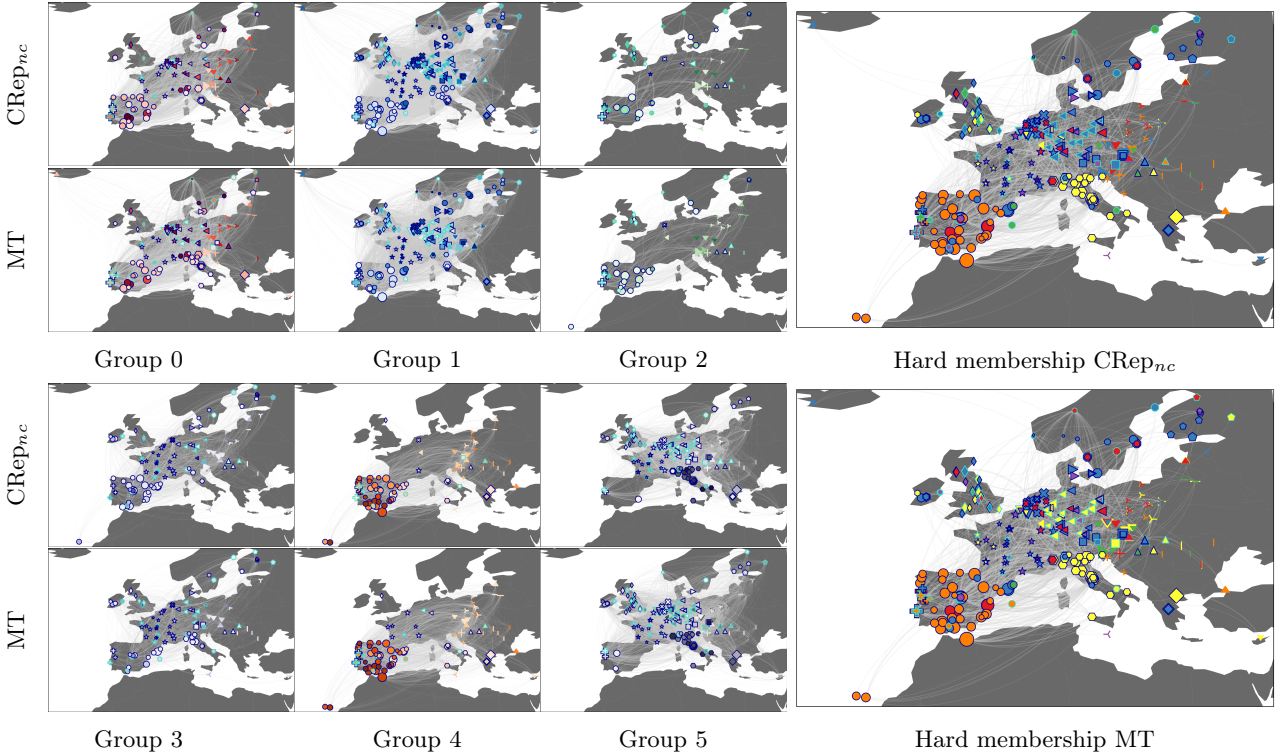
We illustrate our model on a real dataset to show various analysis that a practitioner can perform. We consider a network representation of the Erasmus student exchange program in 2018 [32], denoted as ERS18 in Table S1. A node represents a higher education institution and an edge between nodes  $i$  and  $j$  denotes how many students were sent from  $i$  to spend a portion of their academic year abroad at institution  $j$ , as part of their study program towards a degree (Bachelor, Master, or PhD). This program is supported by the European Commission and involves  $N = 4389$  institutions (mainly European), with a total of  $M = 90972$  participating students in 2018. We recover community partitions from the network data using both CRep<sub>nc</sub> and MT, they have similar and high performance in edge prediction according to AUC (see Table S3), and we fix  $K = 6$  communities from cross-validation. In Fig. 6, we notice that while both models find several groups that closely correlate with countries, CRep<sub>nc</sub> tends to put German institutions (left triangles) more in the same group (blue) and shifts few institutions in the red group, which seems made of mainly universities with strengths in engineering and technology (e.g., Universitat Politècnica de Catalunya, Politecnico di Milano and Institut Polytechnique de Grenoble). For instance, Università di Bologna, Federico II di Napoli and Padova have lower  $u_{i,red}$  than what is predicted without accounting for reciprocity, instead Slovenská technická univerzita v Bratislave, Kauno Technologijos Universitetas and Universidad de Oviedo increase their membership in this group.

In addition, CRep<sub>nc</sub> places more institutions with higher membership in the green group, see Fig. 6 right (hard membership). While there is no apparent common attribute between these (e.g., country), we find that many nodes with high “green” entry of  $u_i$  tend to reciprocate more edges. Specifically, they have a high fraction of out-neighbors such that  $\lambda_{ij}^0$  is much smaller than  $\lambda_{ji}^0$ . That is, the edges  $A_{ij}$  such that  $A_{ji}$  also exists, have a lower impact in determining the value of  $u_i$  in the algorithm. In fact  $u_{ik} \propto \sum_{j,q} A_{ij} \rho_{ij}^{(1)} \phi_{ijkq} = \sum_{j,q} \frac{A_{ij} u_{ik} v_{jq} w_{kq}}{\lambda_{ij}^0 + \eta A_{ji}}$ , see Eq. (7). Hence, if the denominator is high because of  $A_{ji}$ , the weight of the edge  $A_{ij}$  decreases. Nodes with many such  $A_{ij}$  tend to have lower entries  $u_{ik}$  and thus lower  $\lambda_{ij}^0$ . This is a qualitative explanation for having different membership, however the situation is more complicated than this, as one needs to account for the effects on the whole network. In fact, also  $v_{jq}$  changes between the two algorithms, for a similar reason, thus also contributing to a different  $u_{ik}$ .

The primary benefit of CRep, however, lies not in its ability to recover the communities but in what it reveals about the reciprocity patterns in the network. Home and receiving institutions must sign an inter-institutional agreement to allow for student exchanges between them. While institutions may sign them because of clear affinities between their educational training offerings (e.g., both universities are strong in natural science), they might also do so because of some mechanisms involving reciprocity, as hosting students costs resources. Moreover, reciprocity could be further increased by previous knowledge or collaborations between individual faculties, thus institutional reciprocity may be also driven by faculty reciprocity. In addition to the communities themselves, our model also returns  $\eta$ , which can reveal features of the data related to such reciprocity effects not seen with standard generative models, such as  $cr_{ratio}$  or  $\mathbb{E}[A_{ij}|A_{ji}, \Theta]$ . We find a maximum likelihood value of  $\eta = 0.4$ , signaling a significant reciprocity effect. In fact, according to Eq. (16), on average 40% of the edges are influenced by reciprocity.

While  $\eta$  gives a global picture of the whole network, our models still allows to distinguish the impact of reciprocity on individual edges. For instance, if an institution  $i$  accepts many students from  $j$ , then  $j$  might be more willing to accept students from  $i$ , even though  $i$ ’s features might not match  $j$ ’s preferences. If we distinguish the  $u_i$  as the set of preferences of  $i$  and  $v_j$  as the set of attributes of  $j$ , then our model will naturally convey this through high  $\lambda_{ij}^0$  and low  $\lambda_{ji}^0$  for such a case. CRep is able to capture these situations quantitatively, by means of the quantities  $cr_{ij} = \lambda_{ij}^0 / m_{ij}$  (a  $cr_{ratio}$  per edge) with values in  $[0, 1]$  which measures the relative contribution of communities alone to determine edges between  $i$  and  $j$ . Focusing on a single institution  $i$ , one can analyze the difference  $d_{ij} = cr_{ij} - cr_{ji} \in [-1, 1]$  for all  $j$  such that both  $A_{ij}, A_{ji} > 0$  and find different reciprocity patterns, as we show in Fig. 7. Here we plot three extreme cases





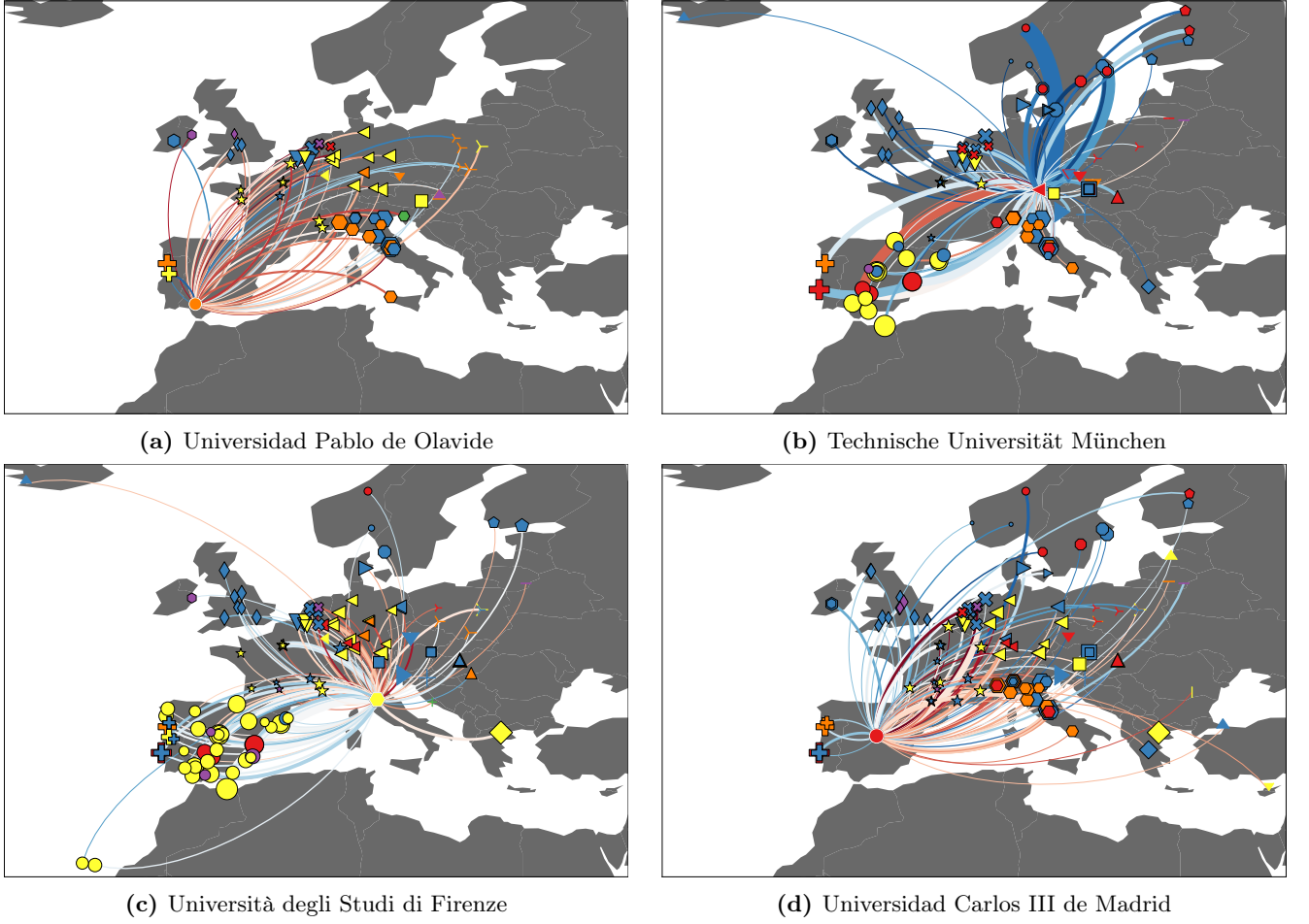
**FIG. 6: Erasmus 2018 community structure.** For visualization clarity, we show the subnetwork made of the 10% biggest institutions and the 3000 edges with highest weights (inference was performed on the whole network). First three columns are one of the  $K = 6$  groups, node color intensity increases with  $u_{ik}$ , so that darker nodes have stronger membership  $u$  in that group, each color is a group (mixed membership) and nodes with light blue border are nodes that change the most the membership in the two algorithms; for each group  $k$ , we only show nodes that have  $u_{ik} > 0.1$ . Rows are  $\text{CRep}_{nc}$  (1st and 3rd) and MT (2nd and 4th). Node and edge size are proportional to the size of an institution measured by the total number of outgoing and incoming students. Node shapes denote country.

where  $i$  has most of the  $d_{ij}$  being less, equal or greater than 0. The Universidad Pablo de Olavide in Sevilla has mostly  $d_{ij} < 0$  (plotted in red), meaning that reciprocity has a strong effect in determining its out-going edges to universities that instead send students to Sevilla mostly out of community preference. The opposite case is that of Technische Universität München, which has most of the  $d_{ij} > 0$  (plotted in blue), signaling that it tends to select its out-going edges more out of preference than their counterparts, who tend to reciprocate instead. Università degli Studi di Firenze is an example of an institution with several  $d_{ij}$  close to 0 (plotted in white), meaning that most of its reciprocated edges are due to community affinities. In other words, Firenze selects out-going  $j$  based on preference and those who select Firenze do the same, so the impact of reciprocity is low. Apart from these three extremes, many universities display a range of such behaviors; we give an example of Universidad Carlos III de Madrid, which has a balanced fraction of reciprocated edges covering these three cases (there are about 1/3 of blue, red, and white edges in the corresponding figure).

## Discussion

CRep is a mathematically principled generative model for capturing both community and reciprocity patterns in directed networks. It relies on relaxing strict conditional independence assumptions on edges that limit the applicability of standard methods on real problems where reciprocity plays an important role. Its algorithmic implementation is efficient and scalable to large system sizes. The corresponding generative model allows for the creation of synthetic networks with the desired interplay between community and reciprocity in determining the edges, while allowing the tuning of network sparsity.

In addition to providing all the analysis tools typical of standard generative models with communities, our model makes it possible to answer questions about reciprocity in networks that were not previously possible, for instance performing probabilistic conditional edge prediction and estimating the relative contribution of community and reciprocity in determining edges. We show how real networks display a wide range of the reciprocity parameter, signaling the variety of possible patterns for this property. In the context of the Erasmus student exchange



**FIG. 7: Patterns of reciprocated edges.** Plots show the subnetwork made of the reciprocated edges of (a) Universidad Pablo de Olavide, (b) Technische Universität München, (c) Università degli Studi di Firenze and (d) Universidad Carlos III de Madrid. Node size is proportional to university size, the shape denotes country, the colors are the highest entry of  $u_i$  (for the four reference nodes - white node border) and  $v_j$  (for all its neighbors). Edge size is proportional to its weight; edge colors vary continuously from red to blue, based on the value of  $d_{ij} = cr_{ij} - cr_{ji}$ : high intensity red, white and high intensity blue mean close to -1, 0 and 1 respectively.

network, our model allowed us to distinguish universities based on their pattern of reciprocated edges.

More generally, our model shows how we can relax strict conditional independence assumptions on edges and showcases possible consequences in doing this. This presents an opportunity for researchers to rethink the fundamental assumptions behind generative models, and present models that may open doors to new theories and questions. We make one step in this direction, as our model connects two popular problems that are mainly treated independently: the inference of communities in networks and generating directed networks where reciprocity plays a relevant role. We used this connection to obtain networks with community structure and values of reciprocity consistent with those observed in real data.

Both the assumption and the model we have presented are only the first step in a broader line of work that investigates how certain topological properties are reflected in

networks with latent community structure as dominant mechanism in edge formation. There are a number of directions in which this work could be extended. We have considered here a simple way to account for reciprocity and break conditional independence, by considering a unique parameter for the whole network. Our model could be extended to account for node-dependent parameters, where reciprocity varies between individuals. In addition, possible extensions may incorporate extra information such as attribute or signals on nodes [34–38], edges of different types as in multilayer networks [15] and dynamics in time [39–44]. Reciprocity is one of the many effects that could play a role in determining how nodes interact in a network. One could go further than this by considering incorporating quantities that account for triples of individuals, for instance clustering coefficient, transitivity or global centrality measures. This is something that exponential random graphs or stochastic



actor oriented models are capable of [13, 45–48], without including latent community structure but rather fitting network statistics. In probabilistic generative models, this would require further breaking conditional dependencies between edges, potentially increasing the model complexity to encompass more complicated situations. With our work, we made the first step in this direction.

While there is no unique generative model that captures all the possible network properties well, our work illustrates how to target reciprocity. As our original motivation to study this problem came from the realization that standard generative models fail to generate synthetic networks with meaningful values for this property, our work illustrates a way in which latent variable frameworks can be applied more realistically, and provides an example of how network scientists can better align fundamental theories with realistic applications.

## Materials and methods

### Synthetic network generation: numerical implementation

The synthetic networks used in the analysis are of three types and represent different scenarios: networks with community structure only, with reciprocity only and networks with both communities and reciprocity. In order to obtain networks with only a community structure we use a stochastic block model with different values of average degree  $\langle k \rangle$ . We generate networks with  $K = 3$  communities of equal-size unmixed group membership,  $N = 2100$  nodes and an assortative structure ( $w$  has higher diagonal entries) with main probabilities  $p_1 = \langle k \rangle * K/N$  and entries outside the main diagonal equal to  $p_2 = 0.1 * p_1$ . We generate three independent samples for each value of  $\langle k \rangle \in [2, 20]$ . On the other hand, we generate networks influenced by reciprocity only through an implementation of the reciprocity model proposed by Holland and Leinhardt (see Supplementary Information Sec. S2 for details). The input parameter  $\alpha$  can be tuned to obtain different values of network reciprocity and we generate three independent samples for each value of  $\alpha \in [0, 10]$ . We consider  $N = 1000$  nodes and a probability to generate one of the directed-edges equal to  $p = 0.002$ . In order to work with synthetic networks having an intrinsic community structure and a given reciprocity value, we use the benchmark generative model proposed in this paper. We generate networks with  $N = 2100$  nodes and  $K = 3$  communities by varying three different input parameters: the average degree  $\langle k \rangle \in [2, 20]$ , the reciprocity coefficient  $\eta \in [0, 1]$  and the fraction of nodes with mixed membership  $over \in [0, 1]$ . While varying one of the parameter, the others are fixed to  $\langle k \rangle = 20$ ,  $\eta = 0.5$  and  $over = 0$ . In detail, networks are generated in two steps. First, membership vectors  $u$  and  $v$  are generated following an equal-size unmixed group membership and a Dirichlet distribution with parameter  $\alpha = 0.1$  for the entries with mixed membership; and the affinity matrix  $w$  is generated us-

ing an assortative block structures with main probabilities  $p_1 = K/N$  and secondary probabilities  $p_2 = 0.1 * p_1$ . Thus the latent variables  $\Theta = (u, v, w, \eta)$  are fixed. Second, edges are drawn according to the generative model described in the main text. Specifically, for each pair of nodes  $(i, j)$ , i) extract  $A_{ij}$  from a Poisson of mean as in Eq. (12); ii) extract  $A_{ji}$  from a Poisson of mean as in Eq. (3). This procedure results in a directed network with the desired reciprocity and sparsity. We generate three independent networks for each value of the three different input parameters.

### Edge prediction and cross-validation

We perform edge prediction using 5-fold cross-validation. In each realization, we divide the dataset, i.e., the entries  $A_{ij}$  of the adjacency matrix, into five equal groups selected at random. We use four of these groups as a training set, to infer the parameters  $\Theta$ . We then use the fifth group as a test set, evaluating the score for each  $A_{ij}$  in this set, and calculate the AUC value. By varying which group we use as the test set, we get 5 trials per realization. The final AUC is the average over these. To compute the regular AUC we use as score the expected value  $\mathbb{E}_{P(A_{ij}|\Theta)}[A_{ij}] = m_{ij}$  as in Eq. (12); for the *conditional* AUC (AUC-cond), we use as score  $\mathbb{E}_{P(A_{ij}|A_{ji}, \Theta)}[A_{ij}] = \lambda_{ij}^0 + \eta A_{ji}$ , i.e., the expected value over the conditional distribution. Notice that the latter can only be computed for CRep, as for the others  $m_{ij} \equiv \lambda_{ij}^0$ , and thus the two AUC values coincide.

### Inference: numerical implementation

All the generative models require inferring  $K$ , the number of communities. We select this by cross-validation. Specifically, we run several held-out trials as explained above by varying  $K$  and select the value of  $K$  that gives the highest (regular) average AUC on the test sets. We then extract the parameters of each method using their best  $K$ . For MT, BPMF and CRep<sub>0</sub>, we extract the parameters  $u, v, w$ ; in addition, for CRep and CRep<sub>nc</sub>, we extract  $\eta$ . All these algorithms converge to a local optima, as the likelihood landscape is not convex. Hence, we run the algorithm 10 times for different random initializations of the parameters and select the realization that has higher likelihood value.

### Acknowledgements

The authors thank Eleanor Power and Elspeth Ready for useful conversations. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Martina Contisciani. **Funding:** All the authors were supported by the Cyber Valley Research Fund. **Author contributions:** All

authors derived the model, analyzed results, and wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to eval-

uate the conclusions in the paper are present in the paper and/or the Supplementary Materials. An open source implementation of the code is available online at <https://github.com/mcontisc/CRep>.

- 
- [1] S. Wasserman, K. Faust, *et al.*, *Social network analysis: Methods and applications*, vol. 8 (Cambridge university press, 1994).
  - [2] L. D. Molm, The structure of reciprocity. *Social psychology quarterly* **73**, 119–131 (2010).
  - [3] M. A. Nowak, K. Sigmund, Evolution of indirect reciprocity. *Nature* **437**, 1291–1298 (2005).
  - [4] D. Garlaschelli, M. I. Loffredo, Structure and evolution of the world trade network. *Physica A: Statistical Mechanics and its Applications* **355**, 138–144 (2005).
  - [5] K. Zhao, X. Wang, M. Yu, B. Gao, User recommendations in reciprocal and bipartite social networks—an online dating case study. *IEEE intelligent systems* **29**, 27–35 (2013).
  - [6] J. Wincent, S. Anokhin, D. Örtqvist, E. Autio, Quality meets structure: Generalized reciprocity and firm-level advantage in strategic networks. *Journal of Management Studies* **47**, 597–624 (2010).
  - [7] W. Li, T. Aste, F. Caccioli, G. Livan, Reciprocity and impact in academic careers. *EPJ Data Science* **8**, 20 (2019).
  - [8] M. E. Newman, S. Forrest, J. Balthrop, Email networks and the spread of computer viruses. *Physical Review E* **66**, 035101 (2002).
  - [9] P. W. Holland, K. B. Laskey, S. Leinhardt, Stochastic blockmodels: First steps. *Social networks* **5**, 109–137 (1983).
  - [10] P. W. Holland, S. Leinhardt, An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association* **76**, 33–50 (1981).
  - [11] J. Park, M. E. Newman, Statistical mechanics of networks. *Physical Review E* **70**, 066117 (2004).
  - [12] D. Garlaschelli, M. I. Loffredo, Multispecies grand-canonical models for networks with reciprocity. *Phys. Rev. E* **73**, 015101 (2006).
  - [13] G. Robins, P. Pattison, Y. Kalish, D. Lusher, An introduction to exponential random graph (p\*) models for social networks. *Social networks* **29**, 173–191 (2007).
  - [14] T. Squartini, F. Picciolo, F. Ruzzenenti, D. Garlaschelli, Reciprocity of weighted networks. *Scientific reports* **3**, 1–9 (2013).
  - [15] C. De Bacco, E. A. Power, D. B. Larremore, C. Moore, Community detection, link prediction, and layer interdependence in multilayer networks. *Physical Review E* **95**, 042317 (2017).
  - [16] B. Ball, B. Karrer, M. E. Newman, Efficient and principled method for detecting communities in networks. *Physical Review E* **84**, 036103 (2011).
  - [17] J. Lloyd, P. Orbanz, Z. Ghahramani, D. M. Roy, *Advances in Neural Information Processing Systems* (2012), pp. 998–1006.
  - [18] P. Orbanz, D. M. Roy, Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE transactions on pattern analysis and machine intelligence* **37**, 437–461 (2014).
  - [19] P. Hoff, *Advances in neural information processing systems* (2008), pp. 657–664.
  - [20] D. J. Aldous, Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis* **11**, 581–598 (1981).
  - [21] D. N. Hoover, Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ* **2** (1979).
  - [22] O. Kallenberg, Multivariate sampling and the estimation problem for exchangeable arrays. *Journal of Theoretical Probability* **12**, 859–883 (1999).
  - [23] J. Besag, Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)* **36**, 192–225 (1974).
  - [24] G. I. Allen, Z. Liu, A local poisson graphical model for inferring networks from sequencing data. *IEEE transactions on nanobioscience* **12**, 189–198 (2013).
  - [25] F. Hadiji, A. Molina, S. Natarajan, K. Kersting, Poisson dependency networks: Gradient boosted models for multivariate count data. *Machine Learning* **100**, 477–507 (2015).
  - [26] D. Strauss, M. Ikeda, Pseudolikelihood estimation for social networks. *Journal of the American statistical association* **85**, 204–212 (1990).
  - [27] A. A. Amini, A. Chen, P. J. Bickel, E. Levina, *et al.*, Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics* **41**, 2097–2122 (2013).
  - [28] D. Garlaschelli, M. I. Loffredo, Patterns of link reciprocity in directed networks. *Phys. Rev. Lett.* **93**, 268701 (2004).
  - [29] P. Gopalan, J. M. Hofman, D. M. Blei, *UAI* (2015), pp. 326–335.
  - [30] A. Ghasemian, H. Hosseinmardi, A. Galstyan, E. M. Airoldi, A. Clauset, Stacking models for nearly optimal link prediction in complex networks. *Proceedings of the National Academy of Sciences* (2020).
  - [31] C. De Bacco, D. B. Larremore, C. Moore, A physical model for efficient ranking in networks. *Science Advances* **4** (2018).
  - [32] Erasmus mobility statistics, <https://data.europa.eu/euodp/en/data>.
  - [33] P. Block, Reciprocity, transitivity, and the mysterious three-cycle. *Social Networks* **40**, 163–173 (2015).
  - [34] M. Contisciani, E. A. Power, C. De Bacco, Community detection with node attributes in multilayer networks. *Scientific Reports* **10**, 15736 (2020).
  - [35] M. E. Newman, A. Clauset, Structure and inference in annotated networks. *Nature communications* **7**, 1–11 (2016).
  - [36] L. Peel, D. B. Larremore, A. Clauset, The ground truth about metadata and community detection in networks. *Science advances* **3**, e1602548 (2017).
  - [37] T. Hoffmann, L. Peel, R. Lambiotte, N. S. Jones, Community detection in networks without observing edges.

- Science advances* **6**, eaav1478 (2020).
- [38] N. Stanley, T. Bonacci, R. Kwitt, M. Niethammer, P. J. Mucha, Stochastic block models with multiple continuous attributes. *Applied Network Science* **4**, 1–22 (2019).
  - [39] A. Ghasemian, P. Zhang, A. Clauset, C. Moore, L. Peel, Detectability thresholds and optimal algorithms for community structure in dynamic networks. *Physical Review X* **6**, 031005 (2016).
  - [40] C. Blundell, J. Beck, K. A. Heller, *Advances in Neural Information Processing Systems* (2012), pp. 2600–2608.
  - [41] X. Zhang, C. Moore, M. E. Newman, Random graph models for dynamic networks. *The European Physical Journal B* **90**, 200 (2017).
  - [42] S. Linderman, R. Adams, *International Conference on Machine Learning* (2014), pp. 1413–1421.
  - [43] T. P. Peixoto, M. Rosvall, Modelling sequences and temporal networks with dynamic community structures. *Nature communications* **8**, 1–12 (2017).
  - [44] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks. *science* **328**, 876–878 (2010).
  - [45] P. Block, C. Stadtfeld, T. A. Snijders, Forms of dependence: Comparing saoms and ergms from basic principles. *Sociological Methods & Research* **48**, 202–239 (2019).
  - [46] T. A. Snijders, Stochastic actor-oriented models for network change. *Journal of mathematical sociology* **21**, 149–172 (1996).
  - [47] T. A. Snijders, The statistical evaluation of social network dynamics. *Sociological methodology* **31**, 361–395 (2001).
  - [48] T. A. Snijders, P. E. Pattison, G. L. Robins, M. S. Handcock, New specifications for exponential random graph models. *Sociological methodology* **36**, 99–153 (2006).
  - [49] V. A. Traag, L. Waltman, N. J. van Eck, From louvain to leiden: guaranteeing well-connected communities. *Scientific reports* **9**, 1–12 (2019).
  - [50] J. Kunegis, *Proc. Int. Conf. on World Wide Web Companion* (2013), pp. 1343–1350.
  - [51] H. Makse (<https://hmakse.ccny.cuny.edu/>).
  - [52] A. Clauset, S. Arbesman, D. B. Larremore, Systematic inequality and hierarchy in faculty hiring networks. *Science Advances* **1** (2015).
  - [53] J. Leskovec, A. Krevl, SNAP Datasets: Stanford large network dataset collection, <http://snap.stanford.edu/data> (2014).
  - [54] P. Ji, J. Jin, Coauthorship and citation networks for statisticians. *Ann. Appl. Stat.* **10**, 1779–1812 (2016).
  - [55] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, Z. Su, *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08 (ACM, New York, NY, USA, 2008), pp. 990–998.

## Supporting Information (SI)

### S1. Detailed derivations

We derive in detail the equations for inferring the parameters. We first apply a variational approach to make the problem tractable, and then use an expectation-maximization algorithm to derive the equations of the updates.

#### A. Variational approach

We aim at maximizing the log-pseudo-likelihood in Eq. (5). The first step is to facilitate the maximization process of the logarithmic term. We consider a probability distribution  $\rho_{ij}$  over the two competing terms: this is our estimate of the probability that the edges exist due to the contribution of either the community membership or the reciprocity term. Applying Jensen's inequality  $\log \bar{x} \geq \overline{\log x}$ :

$$\begin{aligned} \log \lambda_{ij} &= \log \left( \rho_{ij}^{(1)} \frac{\lambda_{ij}^0}{\rho_{ij}^{(1)}} + \rho_{ij}^{(2)} \frac{\eta A_{ji}}{\rho_{ij}^{(2)}} \right) \geq \rho_{ij}^{(1)} \log \frac{\lambda_{ij}^0}{\rho_{ij}^{(1)}} + \rho_{ij}^{(2)} \log \frac{\eta A_{ji}}{\rho_{ij}^{(2)}} \\ &= \rho_{ij}^{(1)} \log \sum_{k,q} u_{ik} v_{jq} w_{kq} + \rho_{ij}^{(2)} \log \eta A_{ji} - \rho_{ij}^{(1)} \log \rho_{ij}^{(1)} - \rho_{ij}^{(2)} \log \rho_{ij}^{(2)} . \end{aligned}$$

Moreover, this holds with equality when:

$$\rho_{ij}^{(1)} = \frac{\lambda_{ij}^0}{\lambda_{ij}^0 + \eta A_{ji}} \quad \text{and} \quad \rho_{ij}^{(2)} = \frac{\eta A_{ji}}{\lambda_{ij}^0 + \eta A_{ji}} . \quad (\text{S1})$$

Thus maximizing  $L^{ps}(\Theta)$  is equivalent to maximizing:

$$L^{ps}(\Theta, \rho) = \sum_{i,j} \left\{ A_{ij} \left( \rho_{ij}^{(1)} \log \sum_{k,q} u_{ik} v_{jq} w_{kq} + \rho_{ij}^{(2)} \log \eta A_{ji} - \rho_{ij}^{(1)} \log \rho_{ij}^{(1)} - \rho_{ij}^{(2)} \log \rho_{ij}^{(2)} \right) - \sum_{k,q} u_{ik} v_{jq} w_{kq} - \eta A_{ji} \right\} .$$

We apply once more the variational approach to make the sum inside the logarithm tractable. Similarly as before, we introduce a probability distribution  $\phi_{ijkq}$  such that:

$$\log \sum_{k,q} u_{ik} v_{jq} w_{kq} \geq \sum_{k,q} \phi_{ijkq} \log u_{ik} v_{jq} w_{kq} - \sum_{k,q} \phi_{ijkq} \log \phi_{ijkq} . \quad (\text{S2})$$

The equality holds when:

$$\phi_{ijkq} = \frac{u_{ik} v_{jq} w_{kq}}{\sum_{k',q'} u_{ik'} v_{jq'} w_{k'q'}} = \frac{u_{ik} v_{jq} w_{kq}}{\lambda_{ij}^0} . \quad (\text{S3})$$

Thus maximizing  $L^{ps}(\Theta, \rho)$  is equivalent to maximizing:

$$\begin{aligned} L^{ps}(\Theta, \rho, \phi) &= \sum_{i,j} \left\{ A_{ij} \rho_{ij}^{(1)} \left( \sum_{k,q} \phi_{ijkq} \log u_{ik} v_{jq} w_{kq} - \sum_{k,q} \phi_{ijkq} \log \phi_{ijkq} \right) + A_{ij} \rho_{ij}^{(2)} \log \eta A_{ji} \right. \\ &\quad \left. - A_{ij} \left( \rho_{ij}^{(1)} \log \rho_{ij}^{(1)} + \rho_{ij}^{(2)} \log \rho_{ij}^{(2)} \right) - \sum_{k,q} u_{ik} v_{jq} w_{kq} - \eta A_{ji} \right\} . \end{aligned} \quad (\text{S4})$$

with respect to  $\Theta, \rho, \phi$ .

### B. Expectation-Maximization updates

Equations for the updates of each of the parameters can be obtained by taking the derivative of Eq. (S4) with respect to a given parameter and setting it to zero. For instance, for the community affinity matrix we get:

$$w_{kq}^{(t+1)} = \frac{\sum_{i,j} A_{ij} \rho_{ij}^{(1)} \phi_{ijkq}}{\sum_{i,j} u_{ik} v_{jq}} . \quad (\text{S5})$$

Here we show how to enforce constraints like  $\sum_k u_{ik} = 1$ , which is an arbitrary choice that can be easily incorporated into our model. For this it is convenient to write the log-pseudo-likelihood as,

$$L^{ps}(\Theta, \rho, \phi) = F(u_{ik}, v_{jq}, w_{kq}) - \sum_{i,j,k,q} u_{ik} v_{jq} w_{kq} \quad (\text{S6})$$

and then substitute  $w_{kq}$  from Eq. (S5) into Eq. (S6):

$$\begin{aligned} L^{ps}(\Theta, \rho, \phi) &= F(u_{ik}, v_{jq}, w_{kq}) - \sum_{i,j,k,q} u_{ik} v_{jq} \frac{\sum_{i,j} A_{ij} \rho_{ij}^{(1)} \phi_{ijkq}}{\sum_{i,j} u_{ik} v_{jq}} \\ &= F(u_{ik}, v_{jq}, w_{kq}) - \sum_{k,q} \sum_{i,j} u_{ik} v_{jq} \frac{\sum_{i,j} A_{ij} \rho_{ij}^{(1)} \phi_{ijkq}}{\sum_{i,j} u_{ik} v_{jq}} \\ &= F(u_{ik}, v_{jq}, w_{kq}) - \sum_{i,j,k,q} A_{ij} \rho_{ij}^{(1)} \phi_{ijkq} . \end{aligned} \quad (\text{S7})$$

The second term in the above equation does not depend explicitly on  $u_{ik}$  and  $v_{jq}$ . In order to apply the constraint on the maximization, we add Lagrange multipliers  $\gamma_i^u, \gamma_j^v$ :

$$L^{ps}(\Theta, \rho, \phi) = F(u_{ik}, v_{jq}, w_{kq}) - \sum_{k,q} \sum_{i,j} A_{ij} \rho_{ij}^{(1)} \phi_{ijkq} - \gamma_i^u \left( \sum_k u_{ik} - 1 \right) - \gamma_j^v \left( \sum_q v_{jq} - 1 \right) . \quad (\text{S8})$$

The update equation for  $u_{ik}$  is obtained by considering the partial derivative,

$$\frac{\partial L^{ps}}{\partial u_{ik}} = \sum_{j,q} \left( \frac{A_{ij} \rho_{ij}^{(1)} \phi_{ijkq}}{u_{ik}} \right) - \gamma_i^u , \quad (\text{S9})$$

and setting it to zero, which yields:

$$u_{ik}^{(t+1)} = \frac{1}{\gamma_i^u} \sum_{j,q} A_{ij} \rho_{ij}^{(1)} \phi_{ijkq} . \quad (\text{S10})$$

By applying the normalization constraint on the  $u_{ik}$ , i.e.,  $\sum_k u_{ik} = 1$ , and noticing that  $\rho_{ij}^{(1)} \phi_{ijkq} = \frac{u_{ik} v_{jq} w_{kq}}{\lambda_{ij}^0 + \eta A_{ji}}$ , we can find an expression for  $\gamma_i^u$ :

$$\gamma_i^u = \sum_{j,k,q} \frac{A_{ij} u_{ik} v_{jq} w_{kq}}{\lambda_{ij}^0 + \eta A_{ji}} = \sum_j \frac{A_{ij} \lambda_{ij}^0}{\lambda_{ij}^0 + \eta A_{ji}} . \quad (\text{S11})$$

Similarly, we have the following update equation for  $v$ :

$$v_{ik}^{(t+1)} = \frac{1}{\gamma_i^v} \sum_{j,q} A_{ji} \rho_{ji}^{(1)} \phi_{jiqk} , \quad (\text{S12})$$

where

$$\gamma_i^v = \sum_{j,k,q} \frac{A_{ji} u_{jq} v_{ik} w_{qk}}{\lambda_{ji}^0 + \eta A_{ij}} = \sum_j \frac{A_{ji} \lambda_{ji}^0}{\lambda_{ji}^0 + \eta A_{ij}} . \quad (\text{S13})$$

### C. Deriving the expected value of the marginal distribution

$$\mathbb{E}[A_{ij}] = m_{ij} = \sum_{A_{ij}, A_{ji}} A_{ij} P(A_{ij}, A_{ji} | \Theta) \quad (\text{S14})$$

$$= \sum_{A_{ji}} P(A_{ji} | \Theta) \sum_{A_{ij}} A_{ij} P(A_{ij} | A_{ji}, \Theta) \quad (\text{S15})$$

$$= \sum_{A_{ji}} P(A_{ji} | \Theta) [\lambda_{ij}^0 + \eta A_{ji}] \quad (\text{S16})$$

$$= \lambda_{ij}^0 + \eta \sum_{A_{ji}} A_{ji} P(A_{ji} | \Theta) \quad (\text{S17})$$

$$= \lambda_{ij}^0 + \eta m_{ji} \quad (\text{S18})$$

$$= \lambda_{ij}^0 + \eta (\lambda_{ji}^0 + \eta m_{ij}) \quad (\text{S19})$$

Solving for  $m_{ij}$  yields:

$$m_{ij} (1 - \eta^2) = (\lambda_{ij}^0 + \eta \lambda_{ji}^0) \quad , \quad (\text{S20})$$

which implies:

$$m_{ij} = \frac{\lambda_{ij}^0 + \eta \lambda_{ji}^0}{(1 - \eta^2)} \quad . \quad (\text{S21})$$

### D. Expected value of $r_w$

With similar calculations as before we obtain:

$$\mathbb{E}[A_{ij} A_{ji}] = \sum_{A_{ij}, A_{ji}} A_{ij} A_{ji} P(A_{ij}, A_{ji} | \Theta) \quad (\text{S22})$$

$$= \lambda_{ij}^0 m_{ji} + \eta \mathbb{E}[A_{ji}^2] \quad . \quad (\text{S23})$$

To fully determine this expression we need to specify the second moment  $\mathbb{E}[A_{ji}^2]$ . For binary variables, we could assume  $\mathbb{E}[A_{ji}^2] = \mathbb{E}[A_{ji}] = m_{ji}$ , as this is the case for Bernoulli distributions. With this assumption, we obtain  $\mathbb{E}[A_{ij} A_{ji}] = (\lambda_{ij}^0 + \eta) m_{ji}$ . Alternatively, we can assume  $\mathbb{E}[A_{ji}^2] = m_{ji} + m_{ji}^2$  as is the case for the Poisson distribution, and thus obtain  $\mathbb{E}[A_{ij} A_{ji}] = (\lambda_{ij}^0 + \eta) m_{ji} + \eta m_{ji}^2$ . Finally we have:

$$\begin{aligned} \mathbb{E}[r_w] &= \frac{\sum_{i,j} \mathbb{E}[A_{ij} A_{ji}]}{\sum_{i,j} \mathbb{E}[A_{ij}]} = \frac{\sum_{i,j} [(\lambda_{ij}^0 + \eta) m_{ji} + \eta m_{ji}^2]}{\sum_{i,j} m_{ij}} \\ &= \eta + \frac{\sum_{i,j} [\lambda_{ij}^0 m_{ji} + \eta m_{ji}^2]}{\sum_{i,j} m_{ij}} \geq \eta \quad . \end{aligned} \quad (\text{S24})$$

With this assumption, we obtain that the parameter  $\eta$  is a lower bound for the expected value of  $r_w$ . An equivalent expression can be derived for models that assume conditional independence, e.g., our model with  $\eta = 0$ . In this case we get:

$$\mathbb{E}[A_{ij} A_{ji}] = \sum_{A_{ij}} A_{ij} P(A_{ij} | \Theta) \sum_{A_{ji}} A_{ji} P(A_{ji} | \Theta) \quad (\text{S25})$$

$$= m_{ij} m_{ji} \quad , \quad (\text{S26})$$

which yields:

$$\mathbb{E}[r_w] = \frac{\sum_{i,j} \mathbb{E}[A_{ij} A_{ji}]}{\sum_{i,j} \mathbb{E}[A_{ij}]} = \frac{\sum_{i,j} m_{ij} m_{ji}}{\sum_{i,j} m_{ij}} \quad . \quad (\text{S27})$$

## S2. Holland and Leinhardt reciprocity model

The model assumes an unweighted and directed network, i.e., asymmetric adjacency matrix with binary values  $A_{ij} \in \{0, 1\}$ , and the following joint probability:

$$P(A|\theta, \alpha) = \frac{e^{-H(A, \theta, \alpha)}}{Z(\theta, \alpha)} \quad (\text{S28})$$

$$H(A, \theta, \alpha) = \theta \sum_{i < j} (A_{ij} + A_{ji}) - \alpha \sum_{i < j} A_{ij} A_{ji} \quad , \quad (\text{S29})$$

where  $Z(\theta, \alpha) = 1 + 2e^{-\theta} + e^{-2\theta + \alpha}$  is the normalization term. The parameter  $\alpha$  controls the level of reciprocity, it couples the two entries  $A_{ij}$  and  $A_{ji}$  thus making the model not factorized; edges between different pairs  $(i, j)$  are conditionally independent given the parameters. This is one of the few analytically tractable exponential random graph models. Due to this property, we can extract analytical marginal and conditional distributions for a pair of nodes  $(i, j)$ :

$$P(A_{ij}|\theta, \alpha) = \frac{e^{-\theta A_{ij}} + e^{-\theta - A_{ij}(\theta - \alpha)}}{Z(\theta, \alpha)} \quad (\text{S30})$$

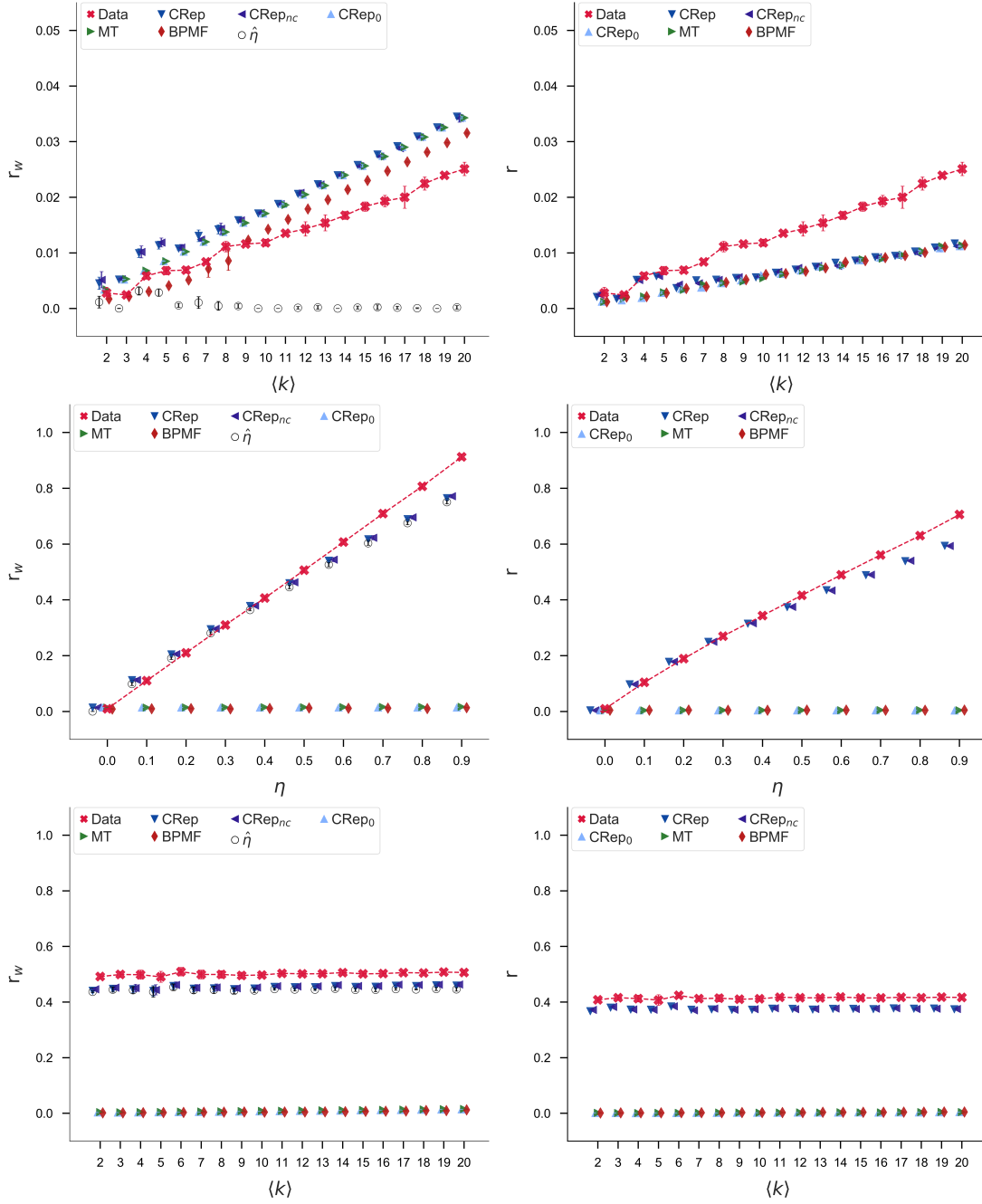
$$P(A_{ji}|A_{ij}, \theta, \alpha) = \frac{e^{-A_{ji}(\theta - \alpha A_{ij})}}{1 + e^{-(\theta - \alpha A_{ij})}} \quad . \quad (\text{S31})$$

These expressions can be used to sample networks with the joint distribution given in Eq. (S29). Tuning the value of the parameter  $\alpha$ , one generates networks with different values of reciprocity.

## S3. Performance in synthetic networks

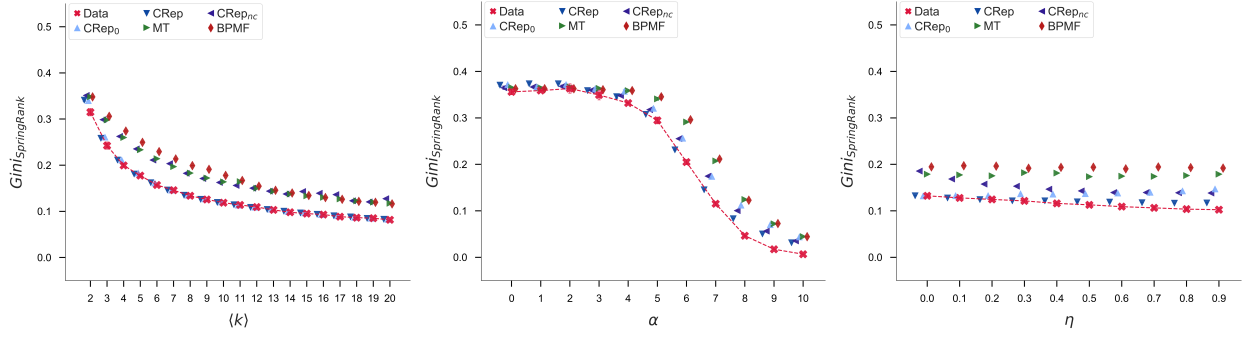
### A. Recovering topological properties

Here we show the ability of the models to reproduce network samples that replicate relevant network quantities. Fig. S1 shows  $r$  and  $r_w$  as defined in Eq. (17), computed in the sampled networks of synthetic data generated with a stochastic block model and our benchmark generative model. As expected, the reciprocity in networks generated with the stochastic block model is always close to zero. Instead, the networks generated with our benchmark generative model present different values of reciprocity, and CRep captures these values significantly better than the other models, consistently across various levels of input  $\eta$  but also for fixed  $\eta$  but different sparsity. By varying the degree of overlapping communities we obtain the same results as varying the average degree (we do not report them here). Fig. S2 shows the Gini index computed on nodes scores obtained with the SpringRank algorithm. The Gini index provides a global measure for the whole network, the higher its value, the more hierarchical the network is. We compare the average over the five samples, and we find that CRep and CRep<sub>0</sub> have reasonable accuracy in retrieving the Gini index of the original network, while the other models tend to overestimate it. This is consistent over the various synthetic network topologies, i.e., network generated with the stochastic block model (left), the HL model (middle) and our benchmark generative model (left). Furthermore, we notice that this topological property decreases as the average degree  $\langle k \rangle$  and  $\alpha$  increase, while it is not influenced by the value of  $\eta$ . We omit the results for the networks generated with our benchmark generative model by varying the sparsity and the fraction of nodes with mixed-membership because we obtain similar results to the stochastic block networks and the benchmark data by varying  $\eta$ , respectively.



**FIG. S1: Reciprocity in synthetic networks.** Synthetic networks with  $N = 2100$  nodes and  $K = 3$  communities of equal-size unmixed group membership generated with a stochastic block model (top) by varying the average degree  $\langle k \rangle$  and our benchmark generative model, by varying the reciprocity parameter  $\eta$  (middle) and  $\langle k \rangle$  (bottom). Results are empirical averages and standard deviations over 15 samples of three independent synthetic networks (5 sample per input network). The red markers indicate the average on the three input networks. (left) The quantity  $r_w$  as defined in Eq. (17);  $\hat{\eta}$  is the inferred parameter in CRep and CRep<sub>nc</sub>. (right) Standard reciprocity  $r$ .

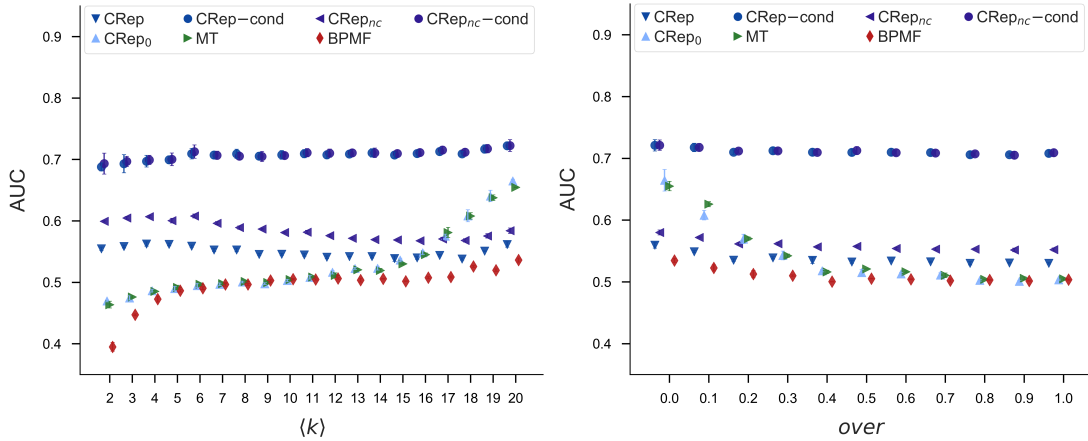




**FIG. S2: Hierarchical structure in synthetic networks.** Synthetic networks generated with (left) the stochastic block model, (middle) the HL model and (right) the benchmark generative model. Results are averages and standard deviations of the Gini index on SpringRank ranking scores over 15 samples of three independent synthetic networks (5 sample per input network). The red markers indicate the average on the three input networks.

### B. Edge prediction in synthetic networks

Here we show the results in terms of edge prediction on synthetic data generated with our benchmark generative model by varying the average degree  $\langle k \rangle$  and the fraction of nodes with mixed-membership, which we denote *over*. We use both conditional and regular edge prediction and Fig. S3 highlights the robustness of CRep and CRep<sub>nc</sub> in terms of conditional edge predictions, as their performance are significantly higher than that of the other algorithms and do not decrease with increasing overlapping communities and sparsity. Notice also the stability of CRep and CRep<sub>nc</sub> in terms of regular edge prediction and how they outperform the other models in critical ranges, e.g., small  $\langle k \rangle$  and high *over*.

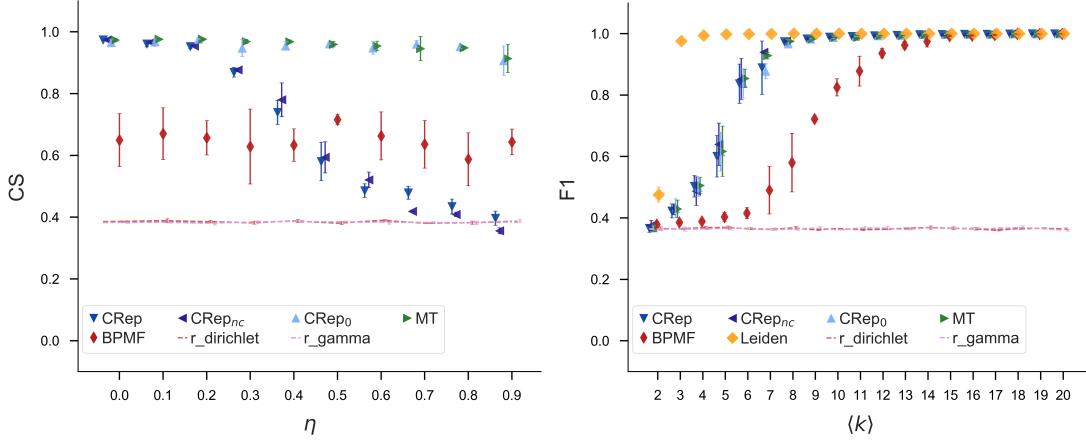


**FIG. S3: Edge prediction in synthetic networks.** Synthetic networks with  $N = 2100$  nodes and  $K = 3$  communities of equal-size unmixed group membership generated with the benchmark generative model proposed above by varying (left) the average degree  $\langle k \rangle$  and (right) the fraction of nodes with mixed-membership *over*. The results are averages and standard deviations over three independent synthetic networks and over 5-folds cross-validation test sets. The accuracy of edge prediction is measured with AUC and the baseline is the random value 0.5.

### C. Community detection in synthetic data

For sake of completeness, here we show the performance of the models on recovering communities. We consider as performance measure the F1-score (F1) and cosine similarity (CS), the former one is valid for hard membership while the latter captures mixed-membership, we calculate for both the average over the nodes. When measuring

the F1-score we consider the entries of maximum value of the membership vectors. Both measures are between 0 and 1 and a value of 1 means perfect reconstruction. Fig. S4 show the accuracy in networks generated with the benchmark generative model by varying the reciprocity parameter  $\eta$  and for synthetic data created with a stochastic block model by varying the average degree  $\langle k \rangle$ . For comparison in these last networks, we consider also the Leiden algorithm [49], a non-generative method. Even if community detection is not the main focus of our model, we notice the ability of CRep in retrieving communities in networks without reciprocity, while its performance decreases as reciprocity increases. This is expected as the community impact in determining the likelihood of an edge decreases as  $\eta$  increases. Notice that the benchmark data have been generated with fixed  $\langle k \rangle = 20$ , thus models without reciprocity are capable of fully recovering the community even in the case where reciprocity is there, provided that the average degree is large enough. These synthetic tests suggest, on one side, the robustness of community detection-only methods in recovering communities even in the presence of reciprocity; on the other side the good performance of CRep in recovering communities when reciprocity has intermediate or low level. This is somehow expected, as this model gives increasingly less weight to the communities as reciprocity increases, thus it is not optimized to recover the communities when these are not fully determining edge formation.



**FIG. S4: Community detection in synthetic networks.** Synthetic networks with  $N = 2100$  nodes and  $K = 3$  communities of equal-size unmixed group membership generated with (left) the benchmark generative model proposed above by varying the reciprocity parameter  $\eta$ , and (right) a stochastic block model. The results are averages and standard deviations over three independent synthetic networks. The accuracy of community detection is measured with (left) cosine similarity and (right) with F1-score as similarity measure, and values close to 1 means higher similarity. The dashed lines represent random baselines, where membership  $u_i$  are extracted randomly from a Dirichlet of parameter  $\alpha = 0.1$  or a Gamma distribution of parameters  $\alpha = 0.1$  and  $\beta = 1$ , to enforce sparsity.

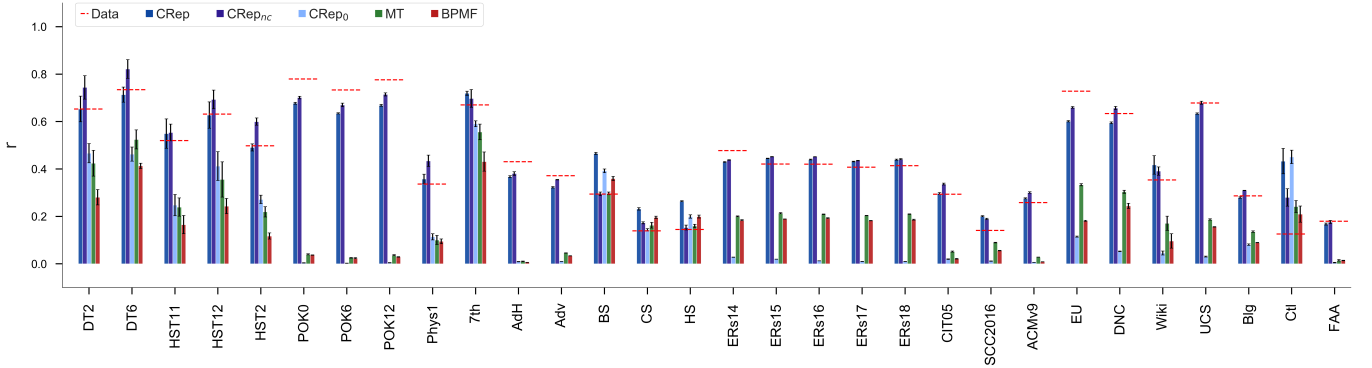
## S4. Performance in real networks

### A. Real data: dataset description

We apply our approach to different types of networks, such as social, infrastructure, online communication, and citation networks. Table S1 provides a brief overview of the datasets studied in this work, as well as their abbreviations. All datasets, have been pre-processed as follows: i) self-loops are removed; ii) only nodes that have at least one out-going and one in-coming edge are kept; iii) we used only the giant connected components. Some datasets require additional specific pre-processing. Specifically, the citation networks (here: CIT05, SCC2016, ACMv9 ) require extracting a network author-author from a network of paper-citation, so that an edge means that an author cites another author. Furthermore, we split dynamic networks into separate individual networks where we kept only interactions happening within a certain time window. This applies to Dutch (DT2, DT6), High school friendships (HST11, HST12, HST2), online dating (POK0, POK6, POK12), and Erasmus (ERs14, ERs15, ERs16, ERs17, ERs18).

### B. Recovering topological properties

Here we show the ability of the models to reproduce network samples that replicate relevant network quantities. For each real network we infer the parameters by each model, and use them to generate five synthetic network samples. Fig. S5 shows the reciprocity  $r$ . For each model, it outputs the averages and the standard deviations over the five samples and the dashed red lines indicate the  $r$  value of the input datasets. We notice the heterogeneity of the analysed networks and how CRep adapts to all different situations, while the other models underestimate the true value most of the time. Fig. S6 shows the Gini index computed on nodes scores obtained with the SpringRank algorithm and the clustering coefficient.



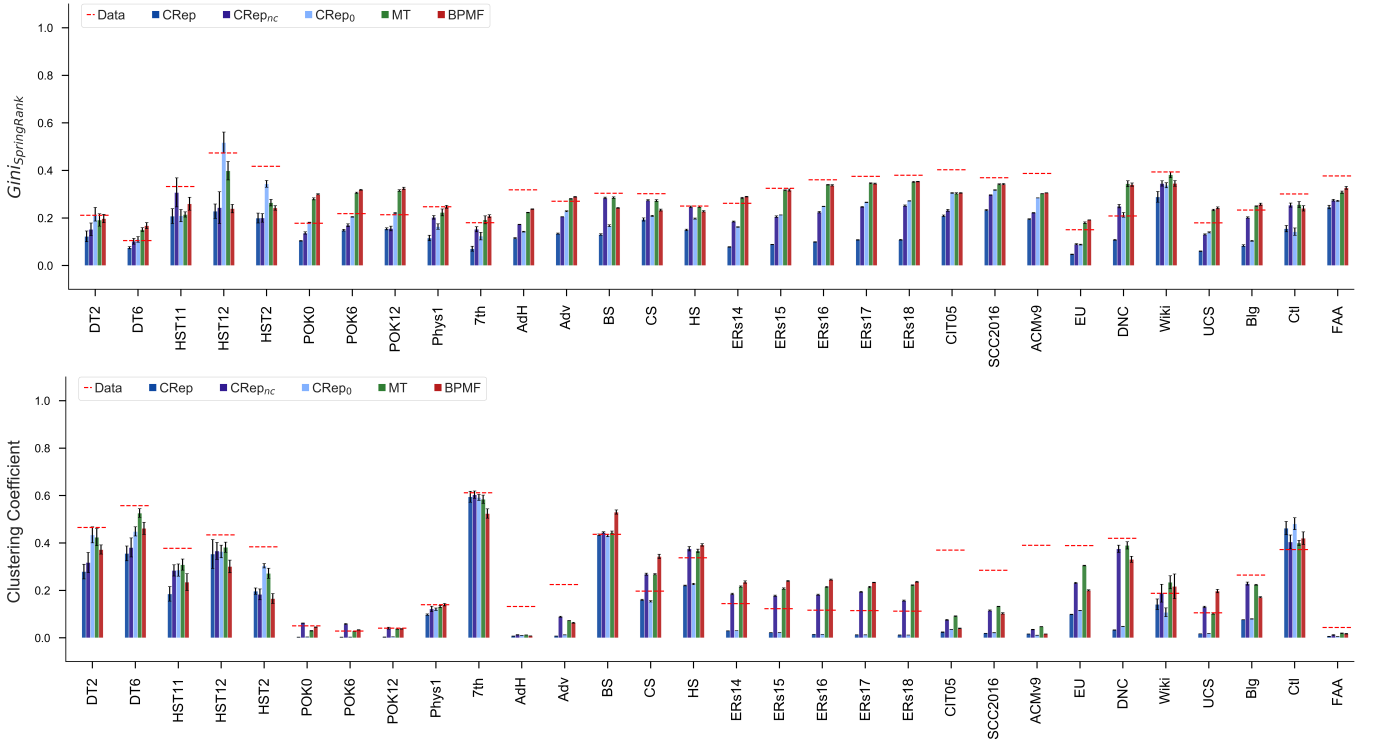
**FIG. S5: Reciprocity in real networks.** Empirical averages and standard deviations of reciprocity  $r$  over 5 samples of each real network (see Table S1 for details). The red dashed lines indicate the  $r$  on the input networks.

### C. Link prediction features

Here we present the supervised learning-link prediction routine (OLP) used for comparison in the edge prediction task on real data. In the link prediction task, scores are assigned to all possible pairs of nodes in the graph based on a set of criteria. Then, the pairs of nodes are sorted according to their scores in an ascending order and the most-likely links are the pairs with scores above a threshold value.

Two categories of features are used to determine the criteria of link classification: (i) global features, defined based on the features of the entire network, such as the number of nodes, number of edges, average degree of nodes, and the average clustering coefficient, and (ii) local features, which include the descriptive features of a single node or a pair of nodes.

In this work, we apply the extended definition of features for a directed network of Ghasemian et al. [30]. We also examine the effect of belonging to the same community on the local pairwise features, i.e., pairwise attributes contribute in the link prediction only if the two nodes belong to the same community. However, we did not find significant changes and at the price of higher computational cost, hence, we exclude this factor from the study and



**FIG. S6: Hierarchical structure and clustering coefficient in real networks.** Empirical averages and standard deviations of (top) the Gini index on SpringRank ranking scores and (bottom) the average clustering coefficient over 5 samples of each real network (see Table S1 for details). The red dashed lines indicate the values on the input networks.

omit the results. Considering  $\Gamma(x)_{out/in}$  as the set of out/in-neighbors of node  $x$ , and  $d(x, y)$  as the distance between nodes  $x$  and  $y$ , some of the well-known features deployed for link prediction are presented in table S2.

**TABLE S1: Datasets description.**

Network	Abbreviation	Category	N	E	Ref.
Dutch college	DT2	Human Social Network	26	144	[50]
Dutch college	DT6	Human Social Network	30	256	[50]
Highschool Friendships	HST11	Human Social Network	31	100	[50]
Highschool Friendships	HST12	Human Social Network	30	114	[50]
Highschool Friendships	HST2	Human Social Network	62	245	[50]
Online dating	POK0	Human Social Network	3562	18098	[51]
Online dating	POK6	Human Social Network	3227	10696	[51]
Online dating	POK12	Human Social Network	2530	7653	[51]
Physicians	Phys	Human Social Network	95	458	[50]
Seventh graders	7th	Human Social Network	29	376	[50]
Adolescent health	AdH	Human Social Network	2213	11676	[50]
Advogato	Adv	Online Social network	3858	42188	[50]
Faculty hiring, business department	BS	Institutions Social Network	112	3321	[52]
Faculty hiring, computer department	CS	Institutions Social Network	198	2702	[52]
Faculty hiring, history department	HS	Institutions Social Network	140	2242	[52]
Erasmus Mobility Statistics 2014	ERs14	Institutions Social Network	2264	79532	[32]
Erasmus Mobility Statistics 2015	ERs15	Institutions Social Network	2890	79665	[32]
Erasmus Mobility Statistics 2016	ERs16	Institutions Social Network	3713	85468	[32]
Erasmus Mobility Statistics 2017	ERs17	Institutions Social Network	4200	89792	[32]
Erasmus Mobility Statistics 2018	ERs18	Institutions Social Network	4389	90972	[32]
Citation 2005	CIT05	Citation Network	2130	11153	[53]
Statistics Citation	SCC2016	Citation Network	2654	21568	[54]
ACM v9 2012	ACMv9	Citation Network	8469	56801	[55]
Email Eu core network	EU	Email Network	834	24348	[53]
DNC Email	DNC	Email Network	548	3575	[50]
Wiki Talk ht	Wiki	Communication Network	80	164	[50]
UC Social	UCS	Communication Network	1302	19044	[50]
Blogs	Blg	Hyperlink Network	830	16107	[50]
Cattle	Ctl	Animal Network	24	191	[50]
FAA Preferred Routes	FAA	Infrastructure Network	1064	2275	[50]

**TABLE S2: Extended features used in the link prediction process for a directed network.**

Feature	Description
Common neighbors out/in	defined for a pair of nodes: $x, y$ : $ \Gamma(x)_{out/in} \cap \Gamma(y)_{out/in} $
Jaccard index	defined for a pair of nodes: $x, y$ : $\frac{ \Gamma(x)_{out/in} \cap \Gamma(y)_{out/in} }{ \Gamma(x)_{out/in} \cup \Gamma(y)_{out/in} }$
Adamic–Adar index	defined for a pair of nodes: $x, y$ : $\sum_{z \in \{\Gamma(x)_{out/in} \cap \Gamma(y)_{out/in}\}} \frac{1}{\log \Gamma(z) }$
Resource Allocation index	defined for a pair of nodes: $x, y$ : $\sum_{z \in \{\Gamma(x)_{out/in} \cap \Gamma(y)_{out/in}\}} \frac{1}{ \Gamma(z) }$
Betweenness centrality	a measure of node centrality based on the shortest paths
Closeness centrality	defined for a pair of nodes: $x, y$ : $\frac{1}{\sum_y d(y, x)}$
Shortest Paths	shortest path between nodes: $x, y$
Katz centralities	a measure of centrality in a network
PageRank centralities	a measure of the importance of a node as an adjustment of Katz centrality
Eigenvector centralities	an adjustment of Katz centrality of a node in regards to the importance of its neighbors
Clustering coefficient for node $x$	$\frac{\text{number of triangles connected to node } x}{\text{number of triples centered around node } x}$
Preferential attachment	the tendency of nodes to connect to the nodes with higher degree
Common community	1 if the pair of nodes belong to the same community, otherwise zero

**TABLE S3: Edge prediction in real networks.** Regular AUC and conditional AUC (AUC–cond) for all real networks (see Table S1 for details). Results are averages and standard deviations over 5-fold cross-validation test sets. In grey box we show the best performance over all methods, while in boldface the best results in terms of regular AUC.

Dataset	AUC						AUC–cond	
	CRep	CRep <sub>nc</sub>	CRep0	MT	BMPF	OLP	CRep	CRep <sub>nc</sub>
DT2	0.71 ± 0.01	<b>0.73 ± 0.01</b>	0.653 ± 0.009	0.71 ± 0.03	0.72 ± 0.01	0.712	0.77 ± 0.02	<b>0.79 ± 0.03</b>
DT6	0.72 ± 0.03	0.76 ± 0.01	0.72 ± 0.01	0.762 ± 0.006	<b>0.774 ± 0.008</b>	0.737	0.83 ± 0.03	<b>0.85 ± 0.02</b>
HST11	<b>0.74 ± 0.01</b>	0.73 ± 0.01	0.63 ± 0.03	0.62 ± 0.03	0.63 ± 0.04	0.714	<b>0.78 ± 0.02</b>	0.76 ± 0.02
HST12	<b>0.82 ± 0.02</b>	0.801 ± 0.008	0.743 ± 0.004	0.74 ± 0.01	0.76 ± 0.02	0.778	0.85 ± 0.01	<b>0.86 ± 0.02</b>
HST2	0.771 ± 0.009	0.76 ± 0.01	0.73 ± 0.01	0.73 ± 0.01	0.71 ± 0.01	<b>0.828</b>	0.808 ± 0.009	0.79 ± 0.02
POK0	0.7747 ± 0.0001	<b>0.845 ± 0.002</b>	0.665 ± 0.002	0.7400 ± 0.0009	0.7652 ± 0.0002	0.804	0.908 ± 0.002	<b>0.934 ± 0.002</b>
POK6	0.758 ± 0.001	<b>0.818 ± 0.002</b>	0.587 ± 0.003	0.626 ± 0.002	0.6939 ± 0.0007	0.750	0.884 ± 0.005	<b>0.909 ± 0.002</b>
POK12	0.765 ± 0.002	<b>0.833 ± 0.002</b>	0.582 ± 0.002	0.606 ± 0.002	0.6723 ± 0.0006	0.739	0.905 ± 0.003	<b>0.924 ± 0.002</b>
Phys1	0.600 ± 0.008	<b>0.627 ± 0.006</b>	0.556 ± 0.009	0.57 ± 0.01	0.60 ± 0.02	0.577	0.676 ± 0.005	<b>0.71 ± 0.01</b>
7th	0.69 ± 0.02	0.79 ± 0.01	0.72 ± 0.02	0.800 ± 0.009	<b>0.809 ± 0.005</b>	0.494	0.77 ± 0.01	<b>0.84 ± 0.01</b>
AdH	0.678 ± 0.003	0.696 ± 0.002	0.656 ± 0.002	0.666 ± 0.003	0.627 ± 0.004	<b>0.867</b>	0.760 ± 0.003	0.787 ± 0.001
Adv	0.771 ± 0.002	0.8919 ± 0.0001	0.760 ± 0.003	0.887 ± 0.001	0.8907 ± 0.0005	<b>0.940</b>	0.830 ± 0.002	0.9333 ± 0.0005
BS	0.662 ± 0.004	<b>0.8749 ± 0.0006</b>	0.649 ± 0.004	0.8749 ± 0.0005	0.8746 ± 0.0009	0.711	0.66 ± 0.01	<b>0.8750 ± 0.0006</b>
CS	0.715 ± 0.008	0.829 ± 0.001	0.696 ± 0.005	0.830 ± 0.002	0.838 ± 0.001	<b>0.844</b>	0.709 ± 0.008	0.833 ± 0.001
HS	0.661 ± 0.005	0.866 ± 0.003	0.646 ± 0.003	0.866 ± 0.003	<b>0.872 ± 0.001</b>	0.865	0.654 ± 0.005	0.867 ± 0.003
ERs14	0.754 ± 0.001	<b>0.9157 ± 0.0005</b>	0.696 ± 0.009	0.9115 ± 0.0004	0.9123 ± 0.0003	0.893	0.810 ± 0.001	<b>0.9278 ± 0.0002</b>
ERs15	0.79 ± 0.01	<b>0.9361 ± 0.0002</b>	0.72 ± 0.02	0.9330 ± 0.0002	0.9312 ± 0.0002	0.929	0.82 ± 0.01	<b>0.9454 ± 0.0002</b>
ERs16	0.8057 ± 0.0006	<b>0.9454 ± 0.0002</b>	0.7064 ± 0.0004	0.9402 ± 0.0003	0.9419 ± 0.0001	0.944	0.8346 ± 0.0006	<b>0.9552 ± 0.0002</b>
ERs17	0.822 ± 0.005	0.9484 ± 0.0001	0.734 ± 0.002	0.9433 ± 0.0002	0.9468 ± 0.0002	<b>0.950</b>	0.838 ± 0.005	<b>0.9568 ± 0.0002</b>
ERs18	0.8334 ± 0.0006	0.9501 ± 0.0001	0.732 ± 0.002	0.9444 ± 0.0002	0.9490 ± 0.0002	<b>0.952</b>	0.8476 ± 0.0006	<b>0.9579 ± 0.0001</b>
CIT05	0.910 ± 0.002	0.9189 ± 0.0008	0.901 ± 0.001	0.918 ± 0.001	0.908 ± 0.001	<b>0.954</b>	0.928 ± 0.002	0.9389 ± 0.0008
SCC2016	0.893 ± 0.001	0.923 ± 0.001	0.8938 ± 0.0009	0.925 ± 0.001	0.9211 ± 0.0007	<b>0.946</b>	0.901 ± 0.001	0.925 ± 0.001
ACMv9	0.926 ± 0.001	0.9350 ± 0.0007	0.919 ± 0.001	0.9352 ± 0.0001	0.9254 ± 0.0006	<b>0.968</b>	0.941 ± 0.001	0.9525 ± 0.0007
EU	0.795 ± 0.007	0.9297 ± 0.0004	0.760 ± 0.007	0.9264 ± 0.0008	0.9169 ± 0.0006	<b>0.944</b>	0.926 ± 0.007	<b>0.9619 ± 0.0006</b>
DNC	0.766 ± 0.003	<b>0.929 ± 0.002</b>	0.730 ± 0.001	0.8566 ± 0.0003	0.913 ± 0.001	0.919	0.890 ± 0.006	<b>0.939 ± 0.002</b>
Wiki	0.68 ± 0.02	0.70 ± 0.02	0.63 ± 0.01	0.63 ± 0.02	<b>0.83 ± 0.01</b>	0.801	0.73 ± 0.01	0.76 ± 0.02
UCS	0.754 ± 0.005	<b>0.8762 ± 0.0008</b>	0.717 ± 0.003	0.8558 ± 0.0008	0.844 ± 0.002	0.850	0.904 ± 0.005	<b>0.9530 ± 0.0008</b>
Blg	0.784 ± 0.001	0.9312 ± 0.0001	0.767 ± 0.002	0.9321 ± 0.0003	<b>0.9334 ± 0.0001</b>	0.924	0.824 ± 0.001	<b>0.9463 ± 0.0001</b>
Ctl	0.56 ± 0.03	0.66 ± 0.02	0.57 ± 0.03	0.67 ± 0.02	<b>0.70 ± 0.03</b>	0.574	0.56 ± 0.03	0.66 ± 0.02
FAA	0.576 ± 0.003	0.589 ± 0.002	0.543 ± 0.007	0.535 ± 0.004	0.607 ± 0.003	<b>0.779</b>	0.592 ± 0.002	0.595 ± 0.002