Community detection and reciprocity in networks by jointly modeling pairs of edges

Martina Contisciani,[1,*] Hadiseh Safdari,[1,†] and Caterina De Bacco[1,‡]

[1] *Max Planck Institute for Intelligent Systems, Cyber Valley, Tuebingen 72076, Germany*

We present a probabilistic generative model and an efficient algorithm to both perform community detection and capture reciprocity in networks. Our approach jointly models pairs of edges with exact 2-edge joint distributions. In addition, it provides closed-form analytical expressions for both marginal and conditional distributions. We validate our model on synthetic data in recovering communities, edge prediction tasks, and generating synthetic networks that replicate the reciprocity values observed in real networks. We also highlight these findings on two real datasets that are relevant for social scientists and behavioral ecologists. Our method overcomes the limitations of both standard algorithms and recent models that incorporate reciprocity through a pseudo-likelihood approximation. We provide an open-source implementation of the code online.

I. Introduction

Network models are powerful and flexible tools for representing complex interactions between individual elements in many different fields [7, 16, 28, 29]. For instance, in social support networks, each individual is a person or the representative of a household, and each link, tie or arc represents the presence or intensity of a relationship between two individuals. Understanding what core patterns drive the observed set of interactions is of high relevance for scientists and practitioners willing to fully exploit the increased availability of networked datasets. A popular approach to modeling networks is that of generative models, in particular latent variable models [10]. They are probabilistic models that introduce latent variables to incorporate domain knowledge and capture complex interactions. Of particular interest, is the possibility of recovering clusters of individuals that behave similarly, a problem named community detection [8]. In this framework, the latent variables represent the nodes' community memberships and the structure of interactions between communities, and the aim is to infer those quantities from the data [1, 6].

Despite their flexibility and computational efficiency, these models have a main flaw: they fail in reproducing important structural network properties such as transitivity, reciprocity, or triadic closure [19, 22, 24]. Synthetic networks generated from these models tend to have significantly lower values of these properties than those observed in real networks.

One possible reason of this problem is the common assumption of conditional independence: conditioned on the latent variables, networks edges are independent and the joint probability distribution is factorized accordingly. This means that an interaction from node i to node j is not directly affected by the interaction in the opposite direction, i.e., the edge $j \to i$. In latent variable models with community structure, as the stochastic block model [12] and its variants, these two edges are fully explained by the membership of the two nodes. While this assumption has been used to obtain tractable problems, it can be too restrictive in certain real scenarios where non-trivial interaction patterns are observed. For instance, in social support networks, it is likely that the existence of interactions from individual i to individual j does not depend only on the groups that i and j belong to, but also on the fact that j has already previously helped i. This tendency of forming mutual connections is called reciprocity [27] and it is an important feature in social networks [20], journal citations [15] and email communications [9, 17], to name a few. While exponential random graph models can represent such network properties in some form [13, 18, 21, 25], they do not incorporate a priori latent variables as community membership. In the previous example, incorporating both community structure and the structural property of reciprocity would help us to understand how an individual interacts with others. Hence, there is a need to incorporate both these phenomena within a unique probabilistic framework.

Recently, Safdari et al. [22] tackled this problem by modeling the conditional distribution of *pairs* of edges between the same nodes, an assumption also shared by seminal works [12, 26]. Safdari et al. [22] include both communities and reciprocity effects inside the likelihood distribution of the network. This resulted in networks samples with values of reciprocity more similar to those of real data, and better edge predictions. However, this model relies on a pseudo-likelihood approximation for parameters' inference, as the model only specifies conditional distributions, but not the *joint* distribution of a pair of edges. As a result of this approximation, the model is not robust in community detection

* martina.contisciani@tuebingen.mpg.de
† hadiseh.safdari@tuebingen.mpg.de
‡ caterina.debacco@tuebingen.mpg.de

in the regime where reciprocity plays a role. Peixoto [19] has shown similar results in terms of triadic closure with a model based on Bayesian inference that combines community structure and this network property. This model also assumes conditional independence among edges and models conditional distributions of triadic edges.

Here we propose a model that takes into account community structure and reciprocity by specifying a closed-form joint distribution of a pair of network edges, which does not involve approximations. To estimate the likelihood of network ties, we use a bivariate Bernoulli distribution–special case of the multivariate Bernoulli distribution– where the log odds are linked to community memberships, and pair-interaction variables. Although these patterns are indicative of two distinct mechanisms of network formation, namely, community structure, and reciprocity, it is reasonable to expect that they are related to each other. For instance, i) the preferred connection between nodes of the same community can induce the presence of reciprocated edges involving similar nodes, and ii) the tendency of forming mutual connections can induce the formation of groups of nodes. This conflation means we cannot reliably interpret the underlying mechanisms of network formation merely from the abundance of reciprocated edges or observed community structure in network data.

Our model takes advantage of the useful properties of the bivariate Bernoulli distribution, i.e., the independence and the uncorrelatedness of the component random variables are equivalent and both the marginal and conditional distributions still follow the Bernoulli distribution. Hence, our model has closed-form analytical expressions and enables practitioners to address with more accuracy questions that were not fully captured by standard models; for instance, predicting the joint existence of mutual ties between pairs of nodes. In addition, its algorithmic implementation is efficient and scalable to large system size, as it exploits the sparsity of network datasets, thus allowing its broad applications across disciplines, e.g., citation networks or neuronal networks that consist of several thousand of nodes.

II. The model

The main goal of this work is to develop a probabilistic generative model with latent variables that better captures real scenarios where non-trivial interaction patterns are observed in networks. This is achieved by modeling *jointly* the edges between the same pair of nodes, differently from standard models that assume their conditional independence given the latent variables. Formally, we model the interactions of N individuals as a binary asymmetric matrix A, with entries A_{ij} defining the presence or the absence of connections from node i to node j. Our model considers jointly the pair $A_{(ij)} := (A_{ij}, A_{ji})$ distributed with a bivariate Bernoulli distribution of parameters $\boldsymbol{\Theta}$, which takes values from $(0,0), (0,1), (1,0)$, and $(1,1)$ in the Cartesian product space $\{0,1\}^2 = \{0,1\} \times \{0,1\}$. Its probability density function can be written as

$$\begin{aligned} P(A_{(ij)}|\boldsymbol{\Theta}) &= P(A_{ij}, A_{ji}|\boldsymbol{\Theta}) \\ &= p_{11}^{A_{ij}A_{ji}} p_{10}^{A_{ij}(1-A_{ji})} p_{01}^{(1-A_{ij})A_{ji}} p_{00}^{(1-A_{ij})(1-A_{ji})} \\ &= \frac{\exp\left\{A_{ij}f_{ij} + A_{ji}f_{ji} + A_{ij}A_{ji}J_{(ij)}\right\}}{Z_{(ij)}}, \end{aligned} \tag{1}$$

where $Z_{(ij)}$ is a normalization constant and $p_{00} = 1/Z_{(ij)}$. In addition, $p_{00} + p_{10} + p_{01} + p_{11} = 1$, and

$$f_{ij} = \log\left(\frac{p_{10}}{p_{00}}\right),\ f_{ji} = \log\left(\frac{p_{01}}{p_{00}}\right),\ J_{(ij)} = \log\left(\frac{p_{11}p_{00}}{p_{10}p_{01}}\right)\ . \tag{2}$$

Thus, $P(A_{ij}, A_{ji}|\boldsymbol{\Theta})$ can be viewed as a member of the exponential family, and can be represented in a log-linear formulation as in Equation (1), where f_{ij}, f_{ji}, and $J_{(ij)}$ represent the natural parameters. $J_{(ij)}$ is called cross-product ratio between A_{ij} and A_{ji} and represents the log-odds of the model. Similar to the Ising model [14], if $J_{(ij)} = 0$ then the components of the bivariate Bernoulli random vector (A_{ij}, A_{ji}) are independent, thanks to the equivalence of independence and uncorrelatedness for multivariate Bernoulli distributions [5]. In this case, the resulting model would be equivalent to consider the product of two independent Bernoulli distributions. Another interesting property of the bivariate Bernoulli is that both marginal and conditional distributions are univariate Bernoulli. Thus, our model has closed-form equations for joint, conditional and marginal distributions.

We now assume that a set of latent variables capture hidden patterns of the data. There are many possibilities to add these variables: one could act directly on the marginal or conditional first moments, as well as modelling separately the different $p_{\alpha\beta}$, with $\alpha, \beta \in \{0,1\}$. However, we model the log ratios to ease interpretability and the

analytical computations. Specifically, we assume

$$f_{ij} = \log \lambda_{ij} \tag{3}$$
$$f_{ji} = \log \lambda_{ji} \tag{4}$$
$$J_{(ij)} = \log \eta \ , \tag{5}$$

where

$$\lambda_{ij} = \sum_{k,q=1}^{K} u_{ik} v_{jq} w_{kq} \tag{6}$$

captures mixed-membership community structure as in De Bacco et al. [6] and η is the pair-interaction coefficient. The parameters u_{ik}, v_{jq} are entries of $K-$dimensional vectors $\boldsymbol{u_i}$ and $\boldsymbol{v_i}$, the out-going and in-coming communities respectively; and 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 (homophily). Thus, $\boldsymbol{\Theta} = (u, v, w, \eta)$ are the latent parameters we want to infer. Through Equations (3)–(5) we encode the assumptions that community structure drives the process of edge formation, and the edges of a pair of nodes depend on each other explicitly according to the parameter η. When $J_{(ij)} = 0$, the probability of $A_{(ij)}$ is given by the agreements of the communities of i and j only; while a positive value for the log-odds will boost the chance to observe a tie between them. Conversely, $J_{(ij)} < 0$ decreases the value of p_{11}, the probability that both edges exist. Considering Equation (5): $0 < \eta < 1$ and $\eta > 1$ codify a negative and positive interaction between i and j, respectively. The first lowers the probability of observing both ties $i \to j$ and $j \to i$, while the latter increases it. Finally, $\eta = 1$ implies no interaction between A_{ij} and A_{ji}.

With this model at hand we can estimate observable quantities, valuable for practitioners. For instance, one can ask about the expected value of a given tie in general or conditioned on the existence of the opposite one, quantities defined as:

$$\mathbb{E}\left[A_{ij}\right] = \frac{\lambda_{ij} + \eta \lambda_{ij} \lambda_{ji}}{Z_{(ij)}} \ , \tag{7}$$

$$\mathbb{E}\left[A_{ij} | A_{ji}\right] = \frac{\eta^{A_{ji}} \lambda_{ij}}{\eta^{A_{ji}} \lambda_{ij} + 1} \ , \tag{8}$$

and similar for $\mathbb{E}\left[A_{ji}\right]$ and $\mathbb{E}\left[A_{ji} | A_{ij}\right]$, see Appendix A. With these quantities one can perform edge prediction tasks, which is crucial when we are limited to a subset of the dataset.

III. Inference

We infer the parameters using a maximum likelihood approach. Specifically, we maximize the log-likelihood

$$\mathcal{L}(\boldsymbol{\Theta}) = \sum_{i,j} f_{ij} A_{ij} + \frac{1}{2} \sum_{i,j} J_{(ij)} A_{ij} A_{ji} - \frac{1}{2} \sum_{i,j} \log Z_{(ij)} \tag{9}$$

with respect to $\boldsymbol{\Theta} = (u, v, w, \eta)$. Adopting a variational approach, this is equivalent to maximize

$$\begin{aligned} \mathcal{L}(\rho, \boldsymbol{\Theta}) = \sum_{i,j} \Bigg[& A_{ij} \sum_{k,q} \rho_{ijkq} \log\Big(\frac{u_{ik} v_{jq} w_{kq}}{\rho_{ijkq}}\Big) + \frac{1}{2} A_{ij} A_{ji} \log \eta \\ & - \frac{1}{2} \log\Big(\sum_{k,q} u_{ik} v_{jq} w_{kq} + \sum_{k,q} u_{jk} v_{iq} w_{kq} + \eta \sum_{k,q} u_{ik} v_{jq} w_{kq} \sum_{k,q} u_{jk} v_{iq} w_{kq} + 1 \Big) \Bigg] \ , \end{aligned} \tag{10}$$

where we introduced the variational distribution ρ_{ijkq} over the parameters and used Jensen's inequality. The equivalence holds when

$$\rho_{ijkq} = \frac{u_{ik} v_{jq} w_{kq}}{\sum_{k,q} u_{ik} v_{jq} w_{kq}} \ . \tag{11}$$

We estimate the parameters by using an expectation-maximation (EM) algorithm where at each step one updates ρ using Equation (11) (E-step) and then maximizes $\mathcal{L}(\rho, \boldsymbol{\Theta})$ with respect to $\boldsymbol{\Theta} = (u, v, w, \eta)$ by setting partial derivatives to zero (M-step). This iteration is repeated until the log-likelihood converges. The exact equations for the updates of the parameters are in Appendix A, and the whole routine is described in Algorithm 1. This algorithm is computationally efficient and scalable to large system sizes as it exploits the sparsity of the dataset. Indeed, all the updates involved in the numerator sum over A_{ij}, hence only the non-zero entries count, giving an algorithmic complexity of $O(M\,K^2)$, where $M = \sum_{i,j} A_{ij}$ is the number of ties.

Algorithm 1 JointCRep: EM algorithm

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

Output: membership matrices $u = [u_{ik}]$, $v = [v_{ik}]$; network-affinity matrix $w = [w_{kq}]$; pair interaction parameter η.

Initialize u, v, w, η at random.

Repeat until $\mathcal{L}$ convergences:

1. Calculate ρ (E-step):

$$\rho_{ijkq} = \frac{u_{ik} v_{jq} w_{kq}}{\sum_{k,q} u_{ik} v_{jq} w_{kq}}$$

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

   i) for each pair (i, k) update memberships:

$$u_{ik} = \frac{\sum_{j,q} A_{ij} \rho_{ijkq}}{\sum_j \left[\frac{\sum_q v_{jq} w_{kq} (1+\eta\lambda_{ji})}{\lambda_{ij}+\lambda_{ji}+\eta\lambda_{ij}\lambda_{ji}+1} \right]}$$

$$v_{ik} = \frac{\sum_{j,q} A_{ji} \rho_{jiqk}}{\sum_j \left[\frac{\sum_q u_{jq} w_{qk} (1+\eta\lambda_{ij})}{\lambda_{ij}+\lambda_{ji}+\eta\lambda_{ij}\lambda_{ji}+1} \right]}$$

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

$$w_{kq} = \frac{\sum_{i,j} A_{ij} \rho_{ijkq}}{\sum_{i,j} \left[\frac{u_{ik} v_{jq} (1+\eta\lambda_{ji})}{\lambda_{ij}+\lambda_{ji}+\eta\lambda_{ij}\lambda_{ji}+1} \right]}$$

   iii) update pair-interaction parameter:

$$\eta = \frac{\sum_{i,j} A_{ij} A_{ji}}{\sum_{i,j} \left[\frac{\lambda_{ij}\lambda_{ji}}{\lambda_{ij}+\lambda_{ji}+\eta\lambda_{ij}\lambda_{ji}+1} \right]}$$

Our model (JointCRep) aims to generalize the method presented in Safdari et al. [22] (CRep), which was of inspiration for the latent variables underlying the generative process. We refer to [22] for a detailed explanation of this method and summarize the main similarities and differences among the models in Table I.

IV. Results

In this section, we present the results obtained in synthetic and real networks. For comparison we use CRep, the model that combines communities and reciprocity with a pseudo-likelihood approximation [22], and MT, a community detection-only generative model with a maximum likelihood approach [6]. Even if both of them posit a Poisson likelihood, in this work we use only binary networks for fair comparisons with our model JointCRep.

A. Results on synthetic data

We first validate the performance of the different methods on synthetic data generated with the model proposed in this work. Being a generative model, given as input an initial set of parameters, one can draw a directed network

	CRep	JointCRep
Networks	Weighted	Binary
Likelihood	Poisson	Bivariate Bernoulli
Marginal mean	$\mathbb{E}[A_{ij}] = \frac{\lambda_{ij}+\eta\lambda_{ji}}{1-\eta^2}$	$\mathbb{E}[A_{ij}] = \frac{\lambda_{ij}+\eta\lambda_{ij}\lambda_{ji}}{Z_{(ij)}}$
Conditional mean	$\mathbb{E}[A_{ij}\|A_{ji}] = \lambda_{ij} + \eta A_{ji}$	$\mathbb{E}[A_{ij}\|A_{ji}] = \frac{\eta^{A_{ji}}\lambda_{ij}}{\eta^{A_{ji}}\lambda_{ij}+1}$
Relationship η *vs* r	Linear	Sublinear
Contribution λ *vs* η	Additive	Multiplicative
Contribution r	Non negative	Real
Closed-form marginal	No	Yes
Closed-form conditional	Yes	Yes
Closed-form joint	No	Yes

TABLE I: Properties of CRep and JointCRep models. λ represents the community effect and η is the parameter linked to the reciprocity r.

with a community structure, and a reciprocity value from the expression in Equation (1). The generative process is described in detail in Appendix B. We analyse networks with $N = 1000$ nodes, $K = 2$ overlapping communities, $\langle k \rangle = 20$ average degree and different values of the pair-interaction parameter η such that we obtain networks with reciprocity values (r) in the interval $[0, 0.8]$. We generate 10 random samples for each value of r.

We test the ability of the models to i) recover the communities, ii) perform edge prediction tasks, and iii) generate sample networks that replicate relevant network quantities.

1. *Community detection*

To evaluate the performance of the methods on recovering the communities, we use the cosine similarity (CS), a measure useful to capture mixed-membership communities, as in this case. It ranges from 0 to 1, where 1 means perfect recovery. We calculate the average of the cosine similarities of both membership matrices u and v, and then averaging over the nodes. The results are shown in Figure 1. In the scenarios with low reciprocity values ($\mathsf{r} < 0.4$) all models perform good. However, as r increases, CRep worsens its performance while JointCRep keeps having good results comparable to those of the community-only algorithm, MT. The big drop of CRep is due to the fact that this model gives increasingly less weight to communities as reciprocity increases, as pointed out in Safdari et al. [22]. Conversely, JointCRep is not affected by the different reciprocity values of the data and still performs as good as MT even by adding another parameter to the model.

2. *Edge prediction*

The edge prediction task consists in estimating the existence of an edge by using the inferred parameters. The main quantity used as a score for the estimation of the entries of the adjacency matrix A is the expected value of the marginal distribution. However, our model also provides the conditional distribution; hence its expected value can also be used as a score. The difference lies in the nature of the question we try to answer. We use the marginal distribution to merely predict the existence of an edge, without considering additional information. On the other hand, with the conditional distribution, we ask what is the probability of an edge $i \to j$, conditioned on observing the state of the opposite edge $j \to i$, i.e., knowing if it exists or not. Here, we exploit the presence or the absence of the edge in the opposite direction to better predict each given entry. Furthermore, our model specifies a joint distribution over the edges of a pair of nodes, and this allows us to answer questions more accurately compared to the standard models, which do not specify a joint distribution. For instance, what is the probability of *jointly* observing both edges or even only an edge in one direction while not observing the other in the opposite? Our model directly captures this by specifying $P(A_{ij}, A_{ji}|\boldsymbol{\Theta})$, while others positing a conditional independence assumption can only compute an approximation as $P(A_{ij}|\boldsymbol{\Theta})\,P(A_{ji}|\boldsymbol{\Theta})$.

In our experiments below, we test edge prediction with various scores by using 5-fold cross-validation. Specifically, we divide the dataset into five equal-size groups (folds) and train the models on four of them (training data) for learning the parameters; this contains 80% of the possible pairs of nodes in the network, so that we hide pairs of entries (A_{ij}, A_{ji}) from the training. One then predicts the existence of edges in the held-out group (test set). As

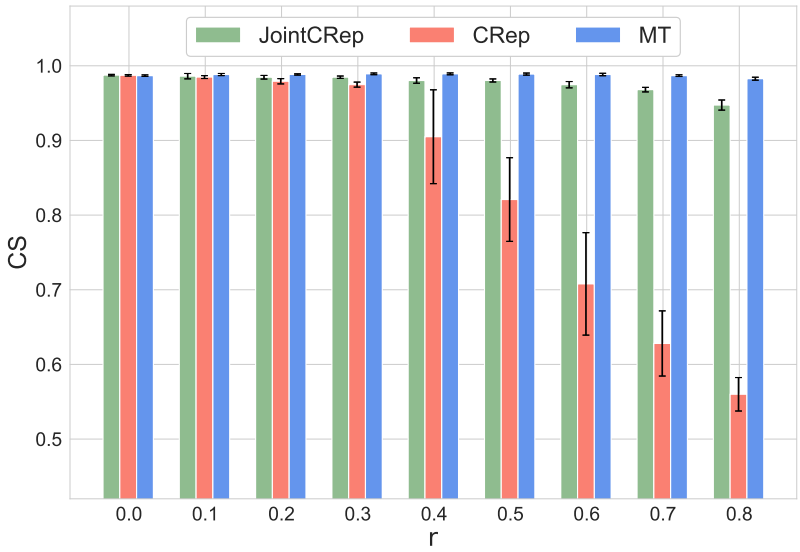

FIG. 1: Community detection in synthetic networks. Cosine similarity (CS) in synthetic networks with $N = 1000$ nodes, $K = 2$ overlapping communities, $\langle k \rangle = 20$ average degree, and different values of reciprocity r. Results are averages and standard deviations over 10 synthetic networks.

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, and conditional AUC values. To estimate the regular AUC, we take the expected value $\mathbb{E}_{P(A_{ij}|\Theta)}[A_{ij}]$ as the score; while for the conditional AUC, the expected value over the conditional distribution, i.e., $\mathbb{E}_{P(A_{ij}|A_{ij},\Theta)}[A_{ij}]$ acts as the score. The latter cannot be computed for the community detection-only algorithm, as the marginal distribution is the same as the conditional, and thus the two AUC values coincide. We provide more details in Appendix C 1, where we also show the ability of our model on edge prediction tasks by using the joint distribution.

Figure 2 displays the results of the marginal and conditional edge prediction for the different models. JointCRep significantly improves the performance of CRep when using the marginal expected value, and it performs as good as MT. The latter, however, is not able to exploit the additional information given by the existence (or non-existence) of the edge in the opposite direction. This dependence is crucial in networks with reciprocity, i.e., most of the real world datasets, and models with an explicit conditional distribution can better adopt this information to obtain higher performance in edge prediction. Indeed, JointCRep and CRep remarkably perform this task, and our model presents more robust results both in terms of standard deviation, and growth.

3. Reproducing the topological properties

A notable property of generative models is their ability to produce synthetic networks based on real-world datasets, such that the synthetic networks imitate the topological features of the real datasets. Following the approach in Safdari et al. [22], for each individual network, we infer the network parameters by applying each model. Then, we use these inferred parameters to generate five network samples. We compare topological properties of these samples with those observed in the ground truth networks used to infer the parameters.

In particular, we are interested in measuring reciprocity. Figure 3 shows the performance of each model in reproducing this feature in sampled networks. As it is expected, MT is not capable of reflecting the observed value of the reciprocity in the ground truth network, a clear indication of the shortcoming of models based purely on community structure, which indeed limits their applications. Conversely, JointCRep perfectly reproduces this quantity. CRep generates sampled networks with reciprocity lower than the ground truth due to the fact that it uses a Poisson likelihood resulting in weighted networks. Additional results are provided in Appendix C 2.

To summarize the results on synthetic networks, JointCRep is capable of recovering communities on networks with varying reciprocity values, performing as good as models that are based purely on community structure. This capability overcomes the limitations of the recent CRep model. Moreover, JointCRep includes many performance enhancements in the edge prediction task, i.e., showing improved results in terms of marginal AUC and more robust conditional AUC values. Furthermore, JointCRep is also capable of generating sampled networks with topological features that resemble that of the real data, e.g., reciprocity and average degree. Collectively, these findings show

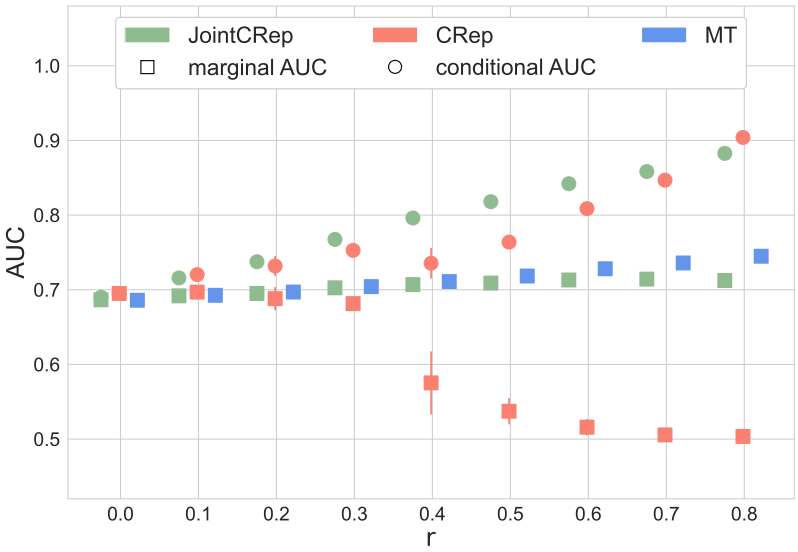

FIG. 2: Edge prediction in synthetic networks. Synthetic networks with $N = 1000$ nodes, $K = 2$ overlapping communities, $\langle k \rangle = 20$ average degree, and different values of reciprocity r. Results are averages and standard deviations over 10 synthetic networks and over 5-folds of cross-validation test sets. Edge prediction performance is measured with AUC and the baseline is the random value 0.5.

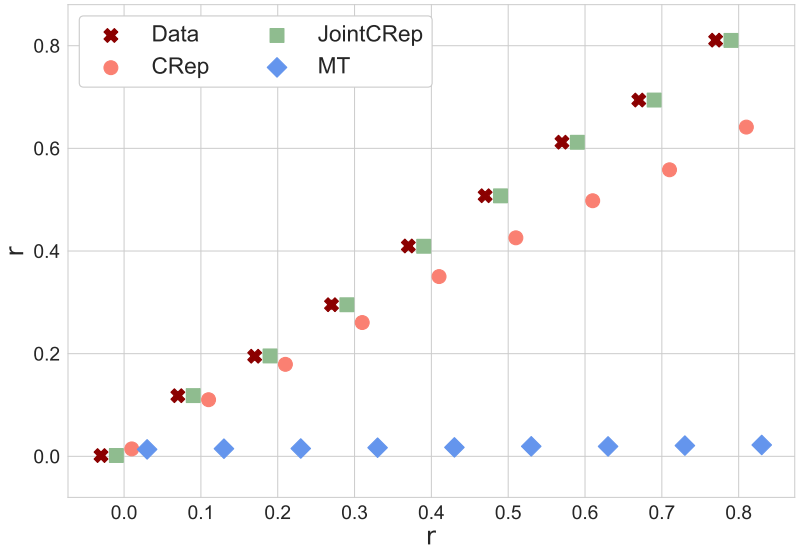

FIG. 3: Reciprocity in sampled synthetic networks. Synthetic networks with $N = 1000$ nodes, $K = 2$ overlapping communities, $\langle k \rangle = 20$ average degree, and different values of reciprocity r. Results are empirical averages and standard deviations over 50 samples of 10 independent synthetic networks (five samples per input network). We measure the reciprocity and the dark red markers indicate the average on 10 input networks.

that JointCRep is able to overcome the limitations of both the community detection-only algorithm MT and the model that incorporates reciprocity through the pseudo-likelihood approximation CRep.

B. Analysis of a high-school social network

We now study the social network that describes friendships between boys in a small high-school in Illinois that was collected in the fall of 1957 [3]. Here, a node represents a boy and an edge from an individual i to j shows that node i claimed to be friend of node j. We pre-processed the dataset by removing self-loops and isolated nodes.

The resulting directed network has 31 nodes, 100 edges and reciprocity equal to 0.52, i.e., only half of the edges (friendship relationships) are reciprocated. There is no additional metadata to describe the nodes, nor is there an available ground truth for the latent parameters. Therefore, we estimate the number of communities K by performing edge prediction task via 5-fold cross-validation with different values of K. For each method the best performance in terms of AUC was achieved with $K = 4$. Figure 4 visualizes the mixed-membership partitions resulting from the matrix u, inferred by the different methods (similar results are obtained for v). Here we use the inferred value of u, which is obtained from the run with the highest log-likelihood over 100 random initializations of the parameters. All the algorithms assign most of the students to the same groups, except from a central block. Here, MT infers mostly hard memberships and balances the number of nodes in each cluster. Instead, CRep allocates only three nodes with small degree to the green community while places the nodes with higher degree in other clusters. JointCRep, shows a partition that lies in between, by inferring mixed-memberships for those nodes known as *bridges*.

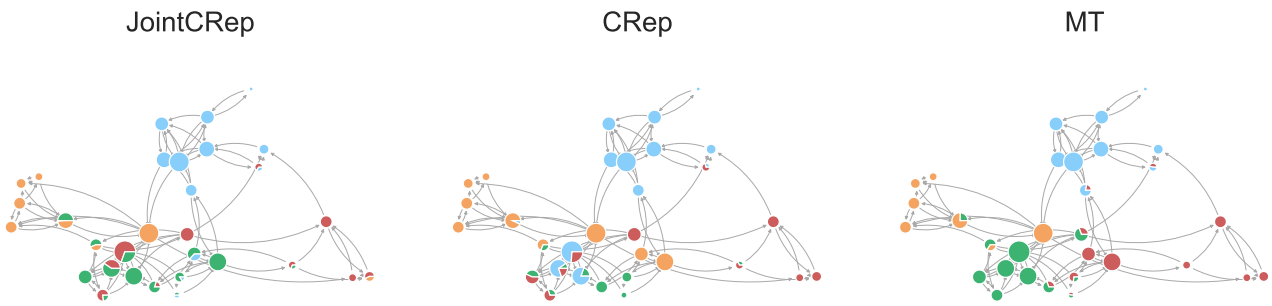

FIG. 4: Community detection in the high-school social network. Mixed-membership partitions determined by the matrix u inferred by JointCRep, CRep, and MT. Node size is proportional to the degree (in- and out-degree).

Given the inferred parameters, we can test the ability of the models to reconstruct the input network, by using either the marginal expected value $\mathbb{E}_{P(A_{ij}|\Theta)}[A_{ij}]$, or the conditional expected value $\mathbb{E}_{P(A_{ij}|A_{ij},\Theta)}[A_{ij}]$ as the score. Note that the latter is not available for MT because the conditional and marginal distributions coincide. Figure 5 presents the results, where edge width and darkness of the reconstructed networks are proportional to the weight given by the expected score (for visualization clarity, we show only edges with weight greater than 0.2). The network estimated by CRep, which uses the expected value of the marginals, does not capture the structure of the data magnificently, as it overestimates the presence of edges. This model specifies conditional distributions and relies on a pseudo-likelihood approximation; since this approach is not necessarily accurate enough to approximate marginals, such results are expected. Instead, MT and JointCRep estimate a sparser representation that is closer to the observed network. However, MT is not able to notably detect reciprocated edges, e.g., $(10, 18)$ or $(64, 67)$, while JointCRep remarkably recovers this type of interactions more precisely. For both JointCRep and CRep, including the conditional expected values improves their accuracy in reconstruction, resulting in identifying reciprocal edges correctly. The difference between the two models lies on the intensity: for instance JointCRep predicts the pair of edges between nodes 10 and 18 with a high probability, while CRep assigns a much lower probability to them. Hence, JointCRep is not only able to predict edges more precisely, but it also does so with higher probability.

To further compare the strength of these methods, we examine their performance in generating samples that resemble the observed network. For each model, we use the inferred parameters to generate five synthetic networks, as shown in Figure 6. Again, we notice how the samples generated by JointCRep better resemble the observed network, as it is easier to distinguish the four blocks generated by JointCRep, compared to the samples from the other algorithms. In particular, JointCRep finds denser groups given by reciprocated edges.

C. Analysis of vampire bat network

As a second example, we study the network of food sharing interactions in captive vampire bats, collected by Carter and Wilkinson [2]. These animals often regurgitate food to roost-mates that fail to feed. The decision of who to feed may depend on both kin relatedness and reciprocal sharing. Hence, in this dataset we expect reciprocity to be an important factor for tie formation. In the study, they fasted 20 vampire bats and induced food sharing on 48 days, over a 2 year period. They showed that reciprocal sharing predicts future food regurgitation more than relatedness or other non-kin food sharing behaviors, such as harassment.

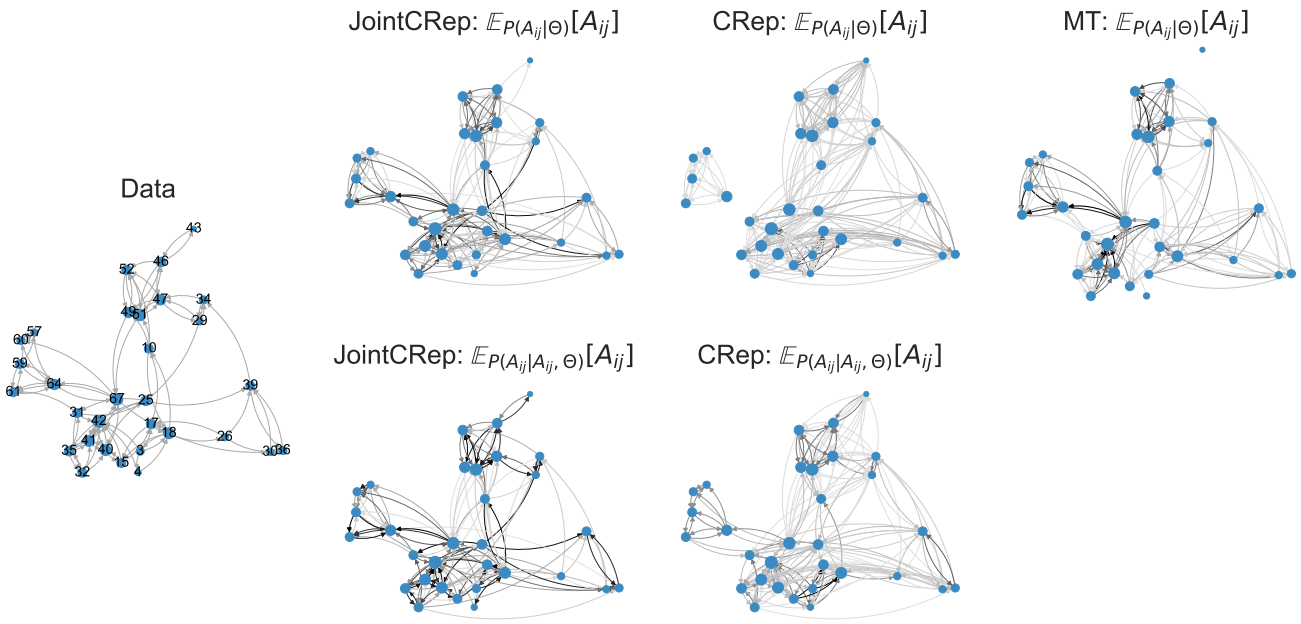

FIG. 5: High-school network reconstruction. (left) High-school data and (right) network reconstructions by using as a score either the marginal expected value $\mathbb{E}_{P(A_{ij}|\Theta)}[A_{ij}]$ or the conditional expected value $\mathbb{E}_{P(A_{ij}|A_{ij},\Theta)}[A_{ij}]$ with the inferred parameters. Note that the last is not available for MT because the conditional and marginal distributions coincide. Edge width and darkness are proportional to the weight (given by the expected score); for visualization clarity we show only edges with weight greater than 0.2. Node size is proportional to the degree (in- and out-degree) and node labels represent node IDs.

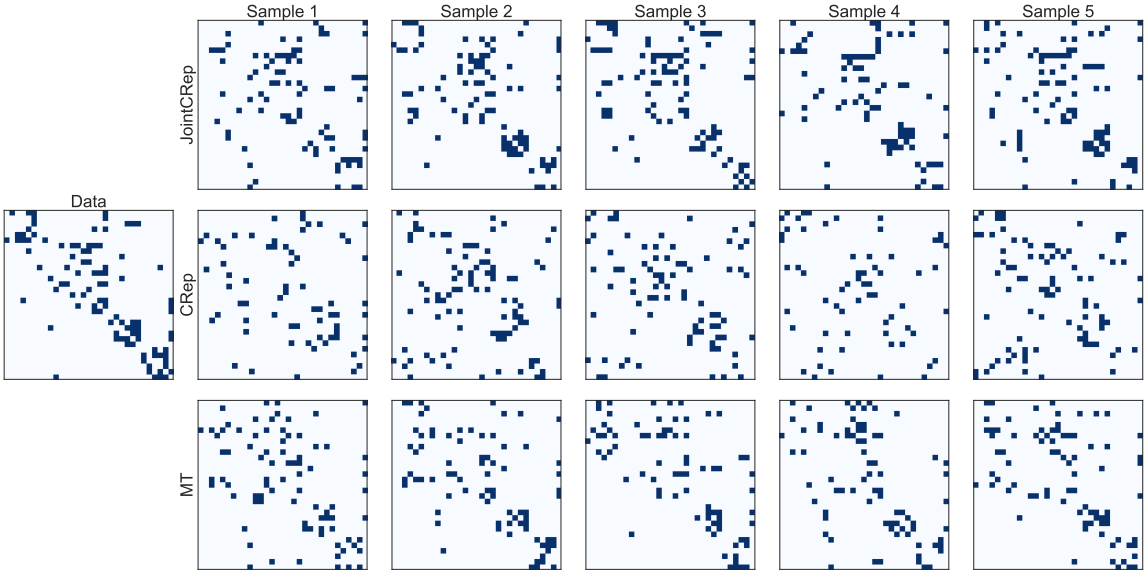

FIG. 6: High-school network samples. (left) High-school data and (right) five random samples generated by different methods with the inferred parameters.

From the collected data, we construct a directed network by building an edge from a bat i to another j if node i fed j at least once. We removed isolated nodes and obtained a network with 19 nodes, 103 edges and reciprocity equal to 0.64. We fix the number of communities $K = 2$ and analyse the data with the different methods. We are interested in measuring the ability of the models to recover the observed network with the inferred parameters, in particular their ability to recover topological properties such as reciprocity. To this aim, we consider the marginal and

the conditional expected values, as in Section IV B. Figure 7 shows the adjacency matrix of the data and its different estimates, obtained by each method. The network embodies a core-periphery structure, where the main core (labels $0-9$) is made of female bats. JointCRep recovers this structure more accurately than other methods, the off-diagonal entries show this fascinating result clearly, while the other methods overestimate the amount of edges. Similarly as observed in the high-school network, our model is not only more accurate, but also assigns higher probabilities to these entries.

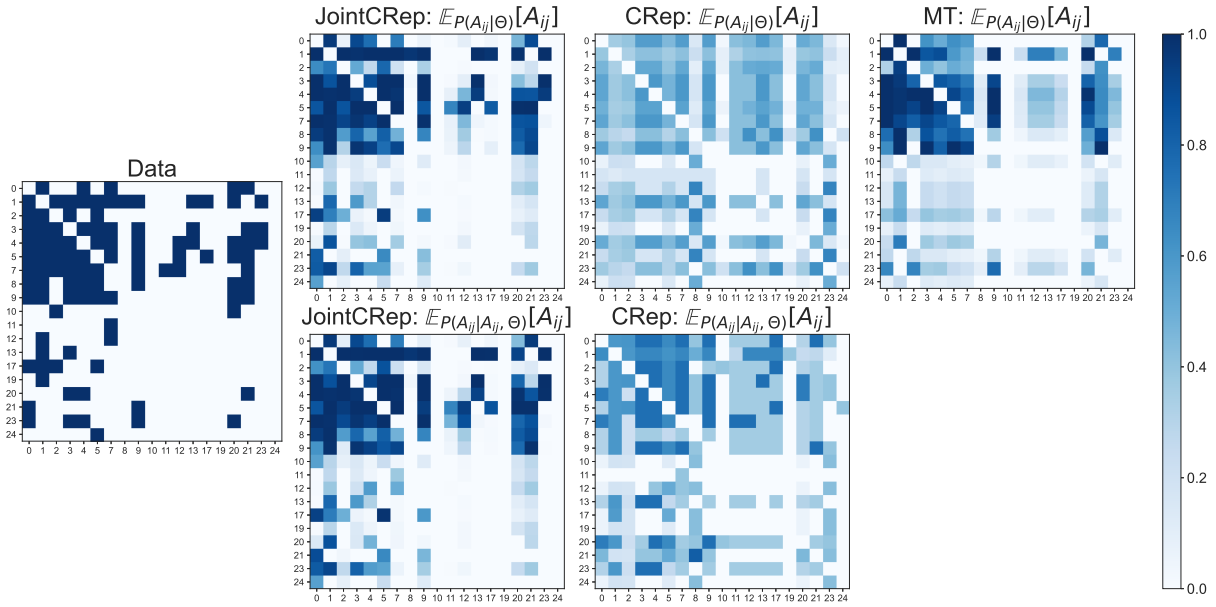

FIG. 7: Vampire bat network reconstruction. (Left) The adjacency matrix of the vampire bat data and (right) its estimates by using as a score either the marginal expected value $\mathbb{E}_{P(A_{ij}|\Theta)}[A_{ij}]$ or the conditional expected value $\mathbb{E}_{P(A_{ij}|A_{ij},\Theta)}[A_{ij}]$ with the inferred parameters. Note that the last is not available for MT because the conditional and marginal distributions coincide. The intensity of the entries is proportional to the score probability, as shown in the colorbar. The labels near the ticks represent node IDs.

In addition to the marginal and conditional expected value, we can consider the joint distribution to estimate the entries of the adjacency matrix. This is equivalent to assign a value to each pair (A_{ij}, A_{ji}) from the set $\{(0,0),(0,1),(1,0),(1,1)\}$, that transforms the edge prediction task into a classification problem. We predict the label associated to the highest probability among $[p_{00}, p_{01}, p_{10}, p_{11}]$, where these are computed by using Equations (S1)–(S4) with the inferred parameters. We assess the goodness of our performance by computing the precision and recall of the predicted labels versus the true labels, as shown in Figure 8. The precision identifies the proportion of correctly classified observed entries. The figure illustrates high precision values consistently across edge labels, as the highest entries are along the diagonal. In particular, JointCRep is able to correctly classify the pairs $(0,0)$ and $(1,1)$. Observing where our model misclassifies, this mainly happens by predicting no edges, i.e., assign label $(0,0)$, when the true ones are either $(0,1)$ or $(1,0)$, implying a tendency to estimate sparser networks. On the other hand, the recall indicates the proportion of predicted edges being correctly classified. Also in this case, the highest entries are in the main diagonal and in predicting the pairs $(0,0)$ and $(1,1)$. Overall, in this case we obtain higher intensities as for the precision, indicating the tendency of labeling the predicted edges with the right type.

To conclude our analysis, we compare five random samples generated with the inferred parameters of each model and check whether they reproduce topological properties as those observed in the real data. Table II shows that JointCRep outperforms other models in terms of all topological properties. In particular, it generates sampled networks with reciprocity values closest to the real data, but also reproduces realistic values of the clustering coefficient.

V. Discussion and conclusion

In this paper, we have presented a generative model called JointCRep that takes into account community structure and reciprocity by specifying a closed-form joint distribution of a pair of network edges, without relying upon approximations. Our model also provides closed-form analytical expressions for both the marginal and conditional

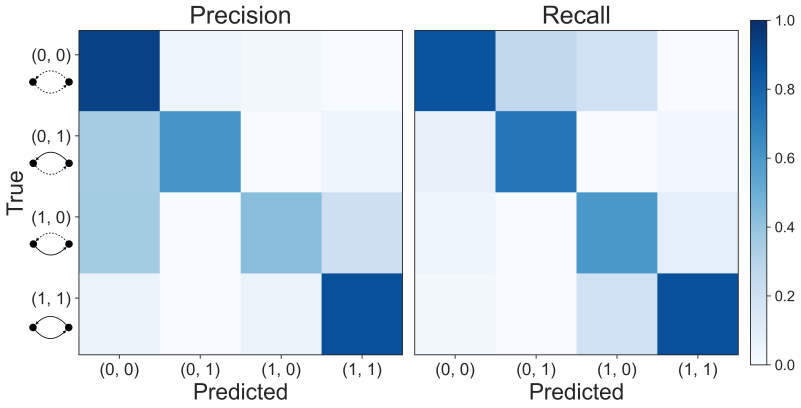

FIG. 8: Precision and recall of the vampire bat network. Statistics based on the confusion matrix that compares the entries of the adjacency matrix and the estimates obtained with the joint distribution of JointCRep. The precision is given by a normalization by row, while the recall accounts for the normalization by column. The label $(0,0)$ denotes no interactions between nodes i and j; labels $(0,1)$ and $(1,0)$ considers the pair of edges where only one edge in one direction is present, and the label $(1,1)$ indicates reciprocated edges.

	N	M	$\langle k \rangle$	r	cc
Data	19	103	10.84	0.64	0.54
JointCRep	18.4 ± 0.89	100.4 ± 5.41	10.92 ± 0.38	0.61 ± 0.03	0.55 ± 0.05
CRep	18.2 ± 0.84	74.2 ± 5.40	8.16 ± 0.54	0.51 ± 0.04	0.27 ± 0.06
MT	17.4 ± 1.14	70 ± 7.38	8.06 ± 0.83	0.36 ± 0.06	0.37 ± 0.01

TABLE II: Topological properties in vampire bat and its sampled networks. Results are averages and standard deviations over five samples. We measure the number of nodes N, the number of edges M, the average degree $\langle k \rangle$, the reciprocity r, and the clustering coefficient cc.

distributions, and enables practitioners to address with more accuracy questions that were not fully captured by standard models; for instance, predicting the joint existence of mutual ties between pairs of nodes.

We first validated our model by applying it to synthetic network datasets, where we achieved remarkable performance in recovering communities, edge prediction tasks, and generating synthetic networks that replicate topological features observed in real networks. We then analyzed two real datasets that are relevant for social scientists and behavioral ecologist, where we found that JointCRep obtains more robust and interpretable results. Collectively, our model is able to overcome the limitations of both standard algorithms and recent models that incorporate reciprocity through the pseudo-likelihood approximation.

The framework we described could be extended in a number of ways. JointCRep analyses binary and single-layer networks; therefore, possible extensions could account for weighted and possibly multilayer networks, where we have edges of different types. Another approach could consider dynamic networks, which have evolving structure over time, and adapt the parameters accordingly [23]. Moreover, our model captures the reciprocity through a unique pair-interaction parameter for the whole network. This model could be improved in the future by including node-dependent parameters in scenarios where reciprocity varies between individuals. Furthermore, many real world datasets contain attributes that provide additional information about their features. Incorporating these extra informations on nodes could result in a more realistic analysis [4].

JointCRep, to the best of our knowledge, is the first such method for fully capturing reciprocity by jointly modeling pairs of edges with exact 2-edge joint distributions. We believe it will serve as a baseline for future models that tackle more complicated interactions that go beyond pairwise interaction, e.g., triadic closure [19].

Acknowledgements

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. **Competing interests**: The authors declare that they have no competing interests.

Data and materials availability:

An open-source algorithmic implementation of the model together with the code to generate synthetic data is publicly available and can be found at https://github.com/mcontisc/JointCRep.

[1] Ball, B., Karrer, B., and Newman, M. E. (2011). Efficient and principled method for detecting communities in networks. *Physical Review E*, 84(3):036103.

[2] Carter, G. G. and Wilkinson, G. S. (2013). Food sharing in vampire bats: reciprocal help predicts donations more than relatedness or harassment. *Proceedings of the Royal Society B: Biological Sciences*, 280(1753):20122573.

[3] Coleman, J. S. (1964). Introduction to mathematical sociology. *London Free Press Glencoe*.

[4] Contisciani, M., Power, E. A., and De Bacco, C. (2020). Community detection with node attributes in multilayer networks. *Scientific reports*, 10(1):1–16.

[5] Dai, B., Ding, S., Wahba, G., et al. (2013). Multivariate bernoulli distribution. *Bernoulli*, 19(4):1465–1483.

[6] De Bacco, C., Power, E. A., Larremore, D. B., and Moore, C. (2017). Community detection, link prediction, and layer interdependence in multilayer networks. *Phys. Rev. E*, 95:042317.

[7] Fell, D. A. and Wagner, A. (2000). The small world of metabolism. *Nature biotechnology*, 18(11):1121–1122.

[8] Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3-5):75–174.

[9] Garlaschelli, D. and Loffredo, M. I. (2004). Patterns of link reciprocity in directed networks. *Physical review letters*, 93(26):268701.

[10] Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233.

[11] Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36.

[12] Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.

[13] Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association*, 76(373):33–50.

[14] Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258.

[15] Li, W., Aste, T., Caccioli, F., and Livan, G. (2019). Reciprocity and impact in academic careers. *EPJ Data Science*, 8(1):20.

[16] Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, 98(2):404–409.

[17] Newman, M. E., Forrest, S., and Balthrop, J. (2002). Email networks and the spread of computer viruses. *Physical Review E*, 66(3):035101.

[18] Park, J. and Newman, M. E. (2004). Statistical mechanics of networks. *Physical Review E*, 70(6):066117.

[19] Peixoto, T. P. (2021). Disentangling homophily, community structure and triadic closure in networks.

[20] Ready, E. and Power, E. A. (2021). Measuring reciprocity: Double sampling, concordance, and network construction. *Network Science*, page 1–16.

[21] Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29(2):173–191.

[22] Safdari, H., Contisciani, M., and De Bacco, C. (2021a). Generative model for reciprocity and community detection in networks. *Phys. Rev. Research*, 3:023209.

[23] Safdari, H., Contisciani, M., and De Bacco, C. (2021b). Reciprocity, community detection, and link prediction in dynamic networks.

[24] Seshadhri, C., Sharma, A., Stolman, A., and Goel, A. (2020). The impossibility of low-rank representations for triangle-rich complex networks. *Proceedings of the National Academy of Sciences*, 117(11):5631–5637.

[25] Snijders, T. A., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological methodology*, 36(1):99–153.

[26] Wasserman, S. and Anderson, C. (1987). Stochastic a posteriori blockmodels: Construction and assessment. *Social networks*, 9(1):1–36.

[27] Wasserman, S., Faust, K., et al. (1994). Social network analysis: Methods and applications.

[28] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442.

[29] Williams, R. J. and Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, 404(6774):180–183.

Supporting Information (SI)

A. Detailed derivations

Combining Equations (2)–(5) we get the explicit mapping between the latent variables and the instances of the joint probability in Equation (1):

$$p_{01} = \frac{\lambda_{ji}}{Z_{(ij)}} \tag{S1}$$

$$p_{10} = \frac{\lambda_{ij}}{Z_{(ij)}} \tag{S2}$$

$$p_{11} = \frac{\eta\lambda_{ij}\lambda_{ji}}{Z_{(ij)}} \tag{S3}$$

$$p_{00} = \frac{1}{Z_{(ij)}} \ , \tag{S4}$$

where the normalization constant is:

$$Z_{(ij)} = \lambda_{ij} + \lambda_{ji} + \eta\lambda_{ij}\lambda_{ji} + 1 \ . \tag{S5}$$

One property of the bivariate Bernoulli is that both marginal and conditional distributions are univariate Bernoulli. Thus, the marginal distributions of A_{ij} and A_{ji} have densities:

$$P(A_{ij}) = (p_{10} + p_{11})^{A_{ij}}(p_{00} + p_{01})^{(1-A_{ij})} \tag{S6}$$

$$P(A_{ji}) = (p_{01} + p_{11})^{A_{ji}}(p_{00} + p_{10})^{(1-A_{ji})} \ , \tag{S7}$$

while the conditional distributions are the following:

$$P(A_{ij}|A_{ji}) = \left(\frac{p(1, A_{ji})}{p(1, A_{ji}) + p(0, A_{ji})}\right)^{A_{ij}} \left(\frac{p(0, A_{ji})}{p(1,_{ji}) + p(0, A_{ji})}\right)^{(1-A_{ij})} \tag{S8}$$

$$P(A_{ji}|A_{ij}) = \left(\frac{p(A_{ij}, 1)}{p(A_{ij}, 1) + p(A_{ij}, 0)}\right)^{A_{ji}} \left(\frac{p(A_{ij}, 0)}{p(A_{ij}, 1) + p(A_{ij}, 0)}\right)^{(1-A_{ji})} \ . \tag{S9}$$

In addition to the expected values reported in the manuscript, we can also compute the variances and the covariance between the random variables:

$$\mathrm{Var}\left[A_{ij}\right] = \left(\frac{\lambda_{ij}(1 + \eta\lambda_{ji})}{Z_{(ij)}}\right)\left(\frac{1 + \lambda_{ji}}{Z_{(ij)}}\right) \tag{S10}$$

$$\mathrm{Var}\left[A_{ji}\right] = \left(\frac{\lambda_{ji}(1 + \eta\lambda_{ij})}{Z_{(ij)}}\right)\left(\frac{1 + \lambda_{ij}}{Z_{(ij)}}\right) \tag{S11}$$

$$\mathrm{Cov}\left[A_{ij}, A_{ji}\right] = \frac{\eta\lambda_{ij}\lambda_{ij} - \lambda_{ij}\lambda_{ij}}{Z^2_{(ij)}} \ . \tag{S12}$$

At each step of the EM algorithm one updates ρ using Equation (11) (E-step) and then maximizes $\mathcal{L}(\rho, \boldsymbol{\Theta})$ with respect to $\boldsymbol{\Theta} = (u, v, w, \eta)$ by setting partial derivatives to zero (M-step). The derivative w.r.t. η is given by:

$$\frac{\partial\mathcal{L}(\rho, \boldsymbol{\Theta})}{\partial\eta} = \frac{1}{2\eta}\sum_{i,j} A_{ij}A_{ji} - \frac{1}{2}\sum_{i,j}\frac{\lambda_{ij}\lambda_{ji}}{\lambda_{ij} + \lambda_{ji} + \eta\lambda_{ij}\lambda_{ji} + 1} \stackrel{!}{=} 0 \quad , \tag{S13}$$

that leads to:

$$\eta = \frac{\sum_{i,j} A_{ij}A_{ji}}{\sum_{i,j}\left[\frac{\lambda_{ij}\lambda_{ji}}{\lambda_{ij} + \lambda_{ji} + \eta\lambda_{ij}\lambda_{ji} + 1}\right]} \quad . \tag{S14}$$

Similarly, we get the updates for u, v and w:

$$u_{ik} = \frac{\sum_{j,q} A_{ij}\rho_{ijkq}}{\sum_j \left[\frac{\sum_q v_{jq}w_{kq}(1+\eta\lambda_{ji})}{\lambda_{ij}+\lambda_{ji}+\eta\lambda_{ij}\lambda_{ji}+1}\right]} \tag{S15}$$

$$v_{ik} = \frac{\sum_{j,q} A_{ji}\rho_{jiqk}}{\sum_j \left[\frac{\sum_q u_{jq}w_{qk}(1+\eta\lambda_{ij})}{\lambda_{ij}+\lambda_{ji}+\eta\lambda_{ij}\lambda_{ji}+1}\right]} \tag{S16}$$

$$w_{kq} = \frac{\sum_{i,j} A_{ij}\rho_{ijkq}}{\sum_{i,j} \left[\frac{u_{ik}v_{jq}(1+\eta\lambda_{ji})}{\lambda_{ij}+\lambda_{ji}+\eta\lambda_{ij}\lambda_{ji}+1}\right]} \ . \tag{S17}$$

B. Benchmark generative model

The model we propose in the manuscript is able to generate synthetic data with intrinsic community structure and a reciprocity value. It takes as input a set of membership vectors, $\boldsymbol{u_i}$ and $\boldsymbol{v_i}$, affinity matrix w, and a pair-interaction parameter η; the output is a directed and binary network with adjacency matrix A whose pairs of edges are conditionally independent from each other. We use the same formulation as in Safdari et al. [22], but our approach differs in that edges between a given pair of nodes are generated stochastically according to the joint probability in Equation (1), and not according to a two-step sampling procedure. In detail, we assign a value to each pair (A_{ij}, A_{ji}) by considering the vector of cumulative probabilities built using Equations (S1)–(S4). To enforce sparsity, we multiply λ by a constant ζ, and in order to automatically rescale the expected value in Equation (7) we have to impose

$$\mathbb{E}[M] = \sum_{i,j} \frac{\zeta\,\lambda_{ij} + \eta\,\zeta\,\lambda_{ij}\,\zeta\,\lambda_{ji}}{\zeta\,\lambda_{ij} + \zeta\,\lambda_{ji} + \eta\,\zeta\,\lambda_{ij}\,\zeta\,\lambda_{ji} + 1} \tag{S1}$$

and solve with respect to ζ, where $\mathbb{E}[M]$ is the expected number of edges, a quantity given in input.

The benchmark we propose here differs from the one presented in Safdari et al. [22] for multiple reasons, as we summarize in Table I. In addition to those, it is worth mentioning that the competing benchmark in Safdari et al. [22] depends on a variable, $cr_{ratio} = 1 - \eta$, that controls the proportion of edges generated purely by either community or reciprocity effect. This implies that in order to have high reciprocity we may weaken the impact of community effect. This does not happen with the model we propose here, as tie formation can be highly influenced by both reciprocity and community structure at the same time, thus providing a more reliable and truthful representation in certain real world examples.

In the manuscript, we use networks generated with the benchmark proposed above where we fix $N = 1000$ nodes, $K = 2$ overlapping communities, $\langle k\rangle = 20$ average degree, and different values of the pair-interaction parameter η such that we obtain networks with reciprocity values r in the interval $[0, 0.8]$. In detail, we use $\eta \in \{0.1, 10, 20, 40, 80, 140, 280, 500, 1500\}$ to get $\mathsf{r} \in \{0, 0.1, 0.2, \ldots, 0.8\}$. To generate the membership matrices u and v we first assign an equal-size unmixed group membership and then we apply the overlapping to 20% of the nodes by drawing those entries from a Dirichlet distribution with parameter $\alpha = 0.1$. The affinity matrix w is generated using an assortative block structure with main probabilities $p_1 = \langle k\rangle\, K \,/\, N$ and secondary probabilities $p_2 = 0.1\, p_1$. Thus the latent variables $\boldsymbol{\Theta} = (u, v, w, \eta)$ are fixed. Then, edges are drawn according to the generative model described above. We generate 10 different samples for each value of η.

For sake of completeness, we also analysed synthetic networks generated with the model proposed in Safdari et al. [22] obtaining similar results and same conclusions. We do not report them here for sake of brevity.

C. Results on synthetic data

1. Edge prediction

We test edge prediction by using a 5-fold cross-validation routine: we divide the dataset into five equal-size groups and train the model on four of them (training set) to infer the parameters; the fifth group is then used as test set to evaluate the existence of edges A_{ij} (in this set). By varying which group we use as test set, we get five trials per realization and the final score is the average over these. To divide the dataset into five folds, we use a symmetric mask, i.e., in each trial the training set contains the 80% of the possible entries (A_{ij}, A_{ji}).

In the manuscript we show the performance of the models in edge prediction when using the marginal and conditional expected values, $\mathbb{E}[A_{ij}]$ and $\mathbb{E}[A_{ij}|A_{ji}]$ respectively. Here, we measure the AUC that is equivalent to the area under the receiver-operating characteristic (ROC) curve [11].

In addition to this results, we can exploit the full joint distribution of our model to answer questions like *what is the probability of jointly observing both edges $i \to j$ and $j \to i$?* This is equivalent to assign a value to the pair (A_{ij}, A_{ji}) from the set $\{(0,0),(0,1),(1,0),(1,1)\}$, that translates the edge prediction task into a classification problem. However, this problem becomes trivial if the model predicts all entries equal to $(0,0)$: in this case we will get high performance just because of the high sparsity of the data. For this reason, we compute the accuracy only for entries in the test set that have at least one edge. For those, we predict the label associated to the highest probability among $[p_{01}, p_{10}, p_{11}]$, where these are computed by using Equations (S1)–(S3) with the inferred parameters. We then compute the accuracy between true and predicted labels, where a value equal to 1 means perfect recovery. As baselines, we use a uniform random probability over the number of possible labels in the training set (RP), and the maximum relative frequency of the label appearing more often in the training set (MRF). The results are shown in Figure S1, where we can observe a V-shape. Reciprocity equal to zero ($\mathsf{r} = 0$) means the networks have no reciprocated edges, and higher its value higher the frequency of the label $(1,1)$. Thus, in the regime $0 \leq \mathsf{r} \leq 0.5$ the performance decreases because the problem becomes more difficult by reaching the point where labels have similar relative frequencies (MRF $\approx$ RP when $\mathsf{r} = 0.5$). In this scenario, JointCRep outperforms the baselines with a bigger gap as the reciprocity increases. When $\mathsf{r} > 0.5$ the problem becomes easier due to the increasing proportion of the label $(1,1)$. Here, predicting all entries equal to $(1,1)$ results in higher performance (MRF). However, this is another trivial situation that should be ignored when analyzing the performance in edge prediction tasks.

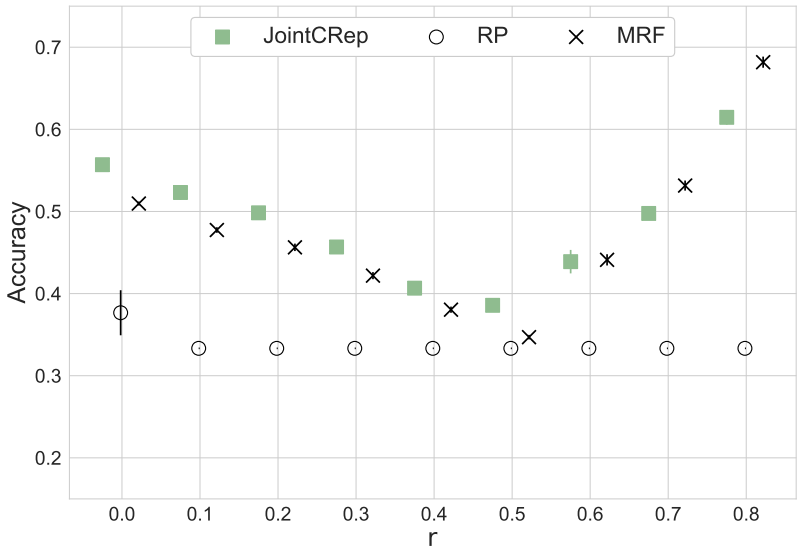

FIG. S1: Edge prediction with joint distributions in synthetic networks. Synthetic networks with $N = 1000$ nodes, $K = 2$ overlapping communities, $\langle k \rangle = 20$ average degree, and different values of reciprocity r. Results are averages and standard deviations over 10 synthetic networks and over 5-folds of cross-validation test sets. Edge prediction performance is measured with accuracy, and as baselines we consider the uniform random probability (RP) and the maximum relative frequency (MRF).

2. Reproducing network topological properties

Figure S2 shows the performance of each model in reproducing the average degree in sampled networks. While JointCRep and MT recover this feature despite the different values of reciprocity, CRep produces samples with a lower average degree than the one given in input as r increases. This happens because, in high reciprocity settings, CRep produces sampled networks with fewer edges but higher weights. Hence, while the average degree decreases, the weighted average degree better reflects the input feature (not shown here).

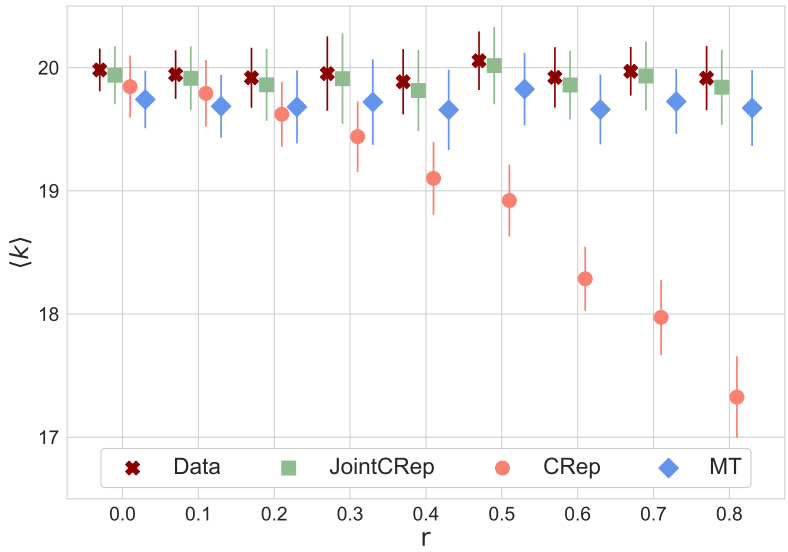

FIG. S2: Average degree in sampled synthetic networks. Synthetic networks with $N = 1000$ nodes, $K = 2$ overlapping communities, $\langle k \rangle = 20$ average degree, and different values of reciprocity r. Results are empirical averages and standard deviations over 50 samples of 10 independent synthetic networks (five samples per input network). We measure the average degree and the dark red markers indicate the average on 10 input networks.