Reducing Degeneracy in Maximum Entropy Models of Networks

Szabolcs Horvát,1 Éva Czabarka,2 and Zoltán Toroczkai1

1Department of Physics, University of Notre Dame, Notre Dame, IN, 46556 USA
2Department of Mathematics, University of South Carolina, Columbia, SC, 29208 USA

Based on Jaynes’s maximum entropy principle, exponential random graphs provide a family of principled models that allow the prediction of network properties as constrained by empirical data. However, their use is often hindered by the degeneracy problem characterized by spontaneous symmetry-breaking, where predictions simply fail. Here we show that degeneracy appears when the corresponding density of states function is not log-concave. We propose a solution to the degeneracy problem for a large class of models by exploiting the nonlinear relationships between the constrained measures to convexify the domain of the density of states. We demonstrate the effectiveness of the method on examples, including on Zachary’s karate club network data.

PACS numbers: 89.75.Hc, 89.70.Cf, 05.20.-y, 87.23.Ge

Our understanding and modeling of complex systems is always based on partial information, limited data and knowledge. The only principled method of predicting properties of a complex system subject to what is known (data and knowledge) is based on the Maximum Entropy Principle of Jaynes [1, 2]. Using this principle, he derived the formalism of statistical mechanics, both classical [1] and the time-dependent quantum density-matrix formalism [2], using Shannon’s information entropy [3]. The method generates a probability distribution $P(\mu)$ over all the possible (micro)states of the system by maximizing the entropy $S[P] = -\sum_\mu P(\mu) \ln P(\mu)$ subject to what is known, the latter expressed as ensemble averages over $P(\mu)$. In this context the given data and the available knowledge act as constraints; restricting the set of candidate states describing the system. $P(\mu)$ is then used via the usual partition function formalism to make unbiased predictions about other observables.

The applicability of Jaynes’s method extends well beyond physics, and in particular, it has been applied in biology [4–6], ecology [7, 8], sociology [9, 10], economics [11], engineering [12, 13], computer science [14], etc. It also received attention within network science [15–21], leading to a class of models known as exponential random graphs (ERG). Despite its popularity, however, this method often presents a fundamental problem, the degeneracy problem, that severely hinders its applicability [18, 19]. When this problem occurs, $P(\mu)$ lacks concentration around the averages of the constrained quantities and the typical microstates do not obey the constraints. In case of ERGs, the generated graphs, for example, may either be very sparse, or very dense, but hardly any will have a density close to that of the data network. Predictions based on such distributions can be significantly off. Two basic questions arise related to the degeneracy problem: 1) Under what conditions it occurs? and 2) How can we eliminate or minimize this problem?

In this Letter we answer both questions and present a solution that significantly reduces degeneracy, then illustrate its effectiveness on concrete examples. We will present our analysis and results using the language of networks and ERG models, however, our findings are generally applicable. Let us consider the set $\mathcal{G}_N$ of all labeled simple graphs (no parallel edges, or self-loops) on $N$ nodes, representing the microstates $\mu \mapsto G$, and an arbitrary set of graph measures, or observables $m(G) = m_1(G), \ldots, m_K(G)$, e.g., the number of edges $m_1$, 2-stars $m_2$, triangles $m_3$, the degree of the 9th node. These measures represent the constraints and we assume that we are given specific values $m^0$, for them (input data). They may come from an empirical network $G^0$, or could represent averages from several empirical datasets. A key assumption in Jaynes’s method is to impose these data at the level of ensemble averages:

$$m^0 = \langle m(G) \rangle = \sum_{G \in \mathcal{G}_N} m(G)P(G),$$

and the goal is to determine the ensemble itself, i.e., the probabilities $P(G)$ for all $G$, as constrained by [1] and normalization: $\sum_{G \in \mathcal{G}_N} P(G) = 1$. Since the number of constraints $K$ is usually small, system [1] is strongly underdetermined, the number of unknowns being $|\mathcal{G}_N| = 2^O(N^2)$. Following Jaynes, the least biased distribution $P(G)$ obeying the constraints is the one that maximizes the entropy $S[P] = -\sum_{G \in \mathcal{G}_N} P(G) \ln P(G)$ subject to [1] and normalization. The method of Lagrange multipliers then yields the family of Gibbs distributions:

$$P(G) = P(G; \beta) = \frac{e^{-\sum_{k=1}^K \beta_k m_k(G)}}{Z(\beta)} = \frac{e^{-\beta \cdot m(G)}}{Z(\beta)},$$

where $Z(\beta) = \sum_{G \in \mathcal{G}_N} e^{\beta \cdot m(G)}$ is the partition function. The $\beta = (\beta_1, \ldots, \beta_K)$ are Lagrange multipliers associated with the constraints $m = (m_1, \ldots, m_K)$, determined from solving system [1] with [2], i.e.,

$$\langle m_k \rangle = \frac{\partial F(\beta)}{\partial \beta_k}$$

where $F(\beta) = -\ln Z(\beta)$ denotes the free energy. The average of some other graph measure $q(G)$ in this en-
semble will be \( \langle q \rangle = \sum_{G \in G_N} q(G) P(G; \beta) \). The distribution \( P(G; \beta) \) defines the corresponding exponential random graph model, hereinafter referred to as the ERG(\( \mathbf{m} \)) model. Eq. (3) admits a maximum likelihood interpretation: its solution is the set of parameters \( \beta \) that maximize the probability \( P(G; \beta) = Z^{-1}(\beta) e^{-\beta \cdot \mathbf{m}} \) of the graph \( G^0 \) for which \( \mathbf{m}(G^0) = \mathbf{m}^0 \). Note that all graphs having the same properties \( \mathbf{m} \) will have the same probability in the ERG(\( \mathbf{m} \)) model.

Since the partition function is determined by the graph measures only, we may write \( Z(\beta) = \sum_{\mathbf{m}} N(\mathbf{m}) e^{-\beta \cdot \mathbf{m}} \), where \( N(\mathbf{m}) \) is a counting function, representing the number of graphs that have the same values for these measures, similar to the density of states function in physics. For example, \( N(m_i, m_\Delta) \) is the number of graphs with \( m_i \) edges and \( m_\Delta \) triangles. Let us denote the domain of \( N \) by \( D = \{ \mathbf{m} \in \mathbb{R}^K \mid N(\mathbf{m}) \geq 1 \} \) and by \( L \) its linear size. Accordingly, the probability that a graph sampled by the ERG(\( \mathbf{m} \)) model will have the given values \( \mathbf{m} \) is:

\[
p(\mathbf{m}; \beta) = \frac{N(\mathbf{m})}{Z(\beta)} e^{-\beta \cdot \mathbf{m}},
\]

and thus write (3) as the mean of \( p(\mathbf{m}; \beta) \):

\[
\langle \mathbf{m} \rangle = \sum_{\mathbf{m}} \mathbf{m} p(\mathbf{m}; \beta).
\]

**Sharp constraints.**—In the above the constraints were imposed at the level of averages. It may happen, however, that some of the data holds for all states of the system, akin to integrals of motion in physics. In network science in this case we restrict ourselves to the largest set of graphs \( \mathcal{G}_N(\mathbf{m}^0) \subseteq \mathcal{G}_N \), all having the same value \( \mathbf{m}^0 \) for those particular measures. We refer to these types of constraints as **sharp constraints**. Examples include the set of all graphs with a given number of edges (the \( G(N, M) \) model), introduced by Erdős and Rényi [22], or those with a given degree sequence [23][24], or with given joint-degree matrix [25]. While sharp constraint problems are mathematically hard in general, counting problems, i.e., computing \( N(\mathbf{m}) \), were shown to be the hardest [26][27].

**The degeneracy problem.**—When solving (3) (or (5)) for \( \beta \) with given \( \langle \mathbf{m} \rangle = \mathbf{m}^0 \) we are fixing the parameters \( \beta(\mathbf{m}^0) \equiv \beta^0 \). It may happen that \( p(\mathbf{m}; \beta^0) \) is multimodal, with probability mass concentrated around two or more disjoint and well separated (by distances comparable to \( L \)) domains in the observables \( \mathbf{m} \), in which case the ERG(\( \mathbf{m} \)) is called degenerate. Alternatively, an ERG(\( \mathbf{m} \)) model is not degenerate if for every \( \beta \) value \( p(\mathbf{m}; \beta) \) is not multimodal. As examples, let us consider the two ERG models, ERG(\( m_i, m_\Delta \)) and ERG(\( m_i, m_\Delta \)), shown in Fig. 1. Figures 1(b), 1(d) show \( p(\mathbf{m}; \beta) \) at parameter values corresponding to averages \( (\langle m_i \rangle, \langle m_\Delta \rangle) \) indicated by the black dots. We see that both models are degenerate: for these input values (or corresponding parameters), the sampled graphs will be either very dense or very sparse, practically none with observable values similar to the input data. This is true even in the case when the averages are realizable by specific graphs (seen more clearly in Fig. 1(d)). Observe that the \( \langle \mathbf{m} \rangle \) averages can come from any point in the convex hull of \( D \) (and only from there). Also note that in both cases \( N(\mathbf{m}) \) itself is unimodal, however, \( p(\mathbf{m}; \beta^0) \) is multimodal [28]. It is important to emphasize that degeneracy is meant in the sense that the graphs sampled by \( p(\mathbf{m}; \beta) \) are coming from probability peaks whose separation is large, comparable to \( L \). Strictly speaking, \( N(\mathbf{m}) \) is a combinatorial function and it may be jagged locally (integer effects). However, samples from local, or nearby peaks are similar, which is fine for modeling purposes, it is not considered degeneracy. For that reason, (keeping the notation) in the remainder we will refer to the **smoothened, continuous** version of \( N(\mathbf{m}) \), preserving only its long-wavelength properties. Degeneracy can be best understood in 1D, \( K = 1 \). Let \( f : [a, b] \to \mathbb{R}^+ \) be a twice differentiable positive function, and let \( g(x) = f(x) e^{-\beta x} \). Since \( g(x) > 0 \), the condition for \( g(x) \) not to be multimodal for any \( \beta \) is that it should not have any minima in \((a, b)\) for any \( \beta \). This is true if in any stationary point \( x_0 \), i.e., with \( g'(x_0) = 0 \), the function \( g \) is concave, \( g''(x_0) < 0 \). For a stationary point \( x_0 \) we have \( \beta = \frac{f'(x_0)}{f(x_0)} \). Com-

![Fig. 1. Degenerate ERG models. Plots are from exact enumeration of all labeled graphs on \( N = 9 \) nodes. (a) The counting function \( N(m_i, m_\Delta) \). Color intensity is proportional to the value of \( N \), white means \( N = 0 \) there. (b) Distribution \( p(m_i; \beta) \) from ERG(\( m_i, m_\Delta \)) at \( \beta^0 = 2.20 \) and \( \beta^0 = -0.313 \), corresponding to the black dot. (c) \( N(m_i, m_\Delta) \). (d) \( p(m_i; \beta) \) from ERG(\( m_i, m_\Delta \)) with \( \beta^0 = 1.24 \) and \( \beta^0 = -0.610 \) from the black dot. Insets show 3D versions of the intensity plots. Note from (4) that the domains of \( N \) and \( p \) always coincide.
putting \( g''(x_0) \) and eliminating \( \beta \) from it using the above, we get \( f''(x_0)f(x_0) < f'(x_0)^2 \). Any \( x_0 \in (a, b) \) can be stationary, since \( f(x_0) > 0 \) and thus the corresponding \( \beta = f'(x_0)/f(x_0) \) always exists to make \( x_0 \) stationary. Thus, \( g(x) \) will be non-degenerate if and only if \( f''(x)f(x) - f'(x)^2 < 0 \) for all \( x \in (a, b) \). This is, however, equivalent to saying that \( f(x) \) is strictly log-concave, i.e., \( \ln f(x) \) is (strictly) concave: \( d^2(\ln f(x))/dx^2 < 0 \) for any \( x \in (a, b) \). For example, Gaussians are log-concave. Generalizing this for arbitrary dimensions, we can announce: \textbf{Theorem:} The ERG \((m)\) is non-degenerate if and only if the density of states \( N(m) \) is strictly log-concave.

The necessary and sufficient conditions for function \( N(m) \) to be log-concave \([30]\) is that \( (i) \) its domain \( D \) is convex and \( (ii) \) if \( (i) \) holds, to satisfy the Prékopa–Leindler type inequality \( N(\lambda m + (1 - \lambda)n) > N(m)^\lambda N(n)^{1 - \lambda} \) for any \( 1 < \lambda < 1 \) and \( m, n \in D \). It is important to note that the theorem above reduces degeneracy to purely graph theoretical properties, it has nothing to do with the Gibbs distributions \([2]\). In two or higher dimensions degeneracy occurs frequently, and the typical approach has been simply to switch to an entirely different set of measures \([31]\). Realistically, however, we might not have other data, or its collection would not be an option; we want to extract the maximum possible information from the available data. Additionally, from a domain expertise point of view, e.g., triangle count is a natural variable for sociologists, as it expresses the level of transitivity, an important measure for social networks; yet the corresponding ERG model is degenerate \([15]\).

\textbf{Solution:—} Here we propose to work still with the same variables \( m \) (same data) as in the degenerate ERG model, however, to consider a \textit{one-to-one} transformation \( m \leftrightarrow \xi = F(m) \) such that the corresponding counting function:

\[
\tilde{N}(\xi) = N(F^{-1}(\xi))
\]

is log-concave. Due to the one-to-one nature, one can still work with or plot the distributions in the same coordinate system \( m \) (see Fig. 2(b)(c)), but the graphs are sampled by the non-degenerate model \( ERG(\xi) = ERG(F(m)) \), with constraints \( \xi^0 = F(m^0) = (\xi) \). There is no recipe for obtaining such transformation in general (it might even not exist, e.g., when \( D \) is not singly connected), however, there is a large class of problems where this can be achieved, to which the degenerate models in the literature belong. This is the case when the convexity condition \( (i) \) is violated. To better understand the nature of the \( F \) function in this situation, let us focus on the 2D case. If \( m_1(G) \) and \( m_2(G) \) were independent, \( D \) would be rectangular and therefore convex. Instead, the shapes of the domains in Fig. 1 indicate that there is a nonlinear confining relationship between the variables, on average. For the \((m_1, m_\nu)\) case it holds that \( m_\nu \sim m_1^2 \) on average (Fig. 2(a), thick orange line). Similarly, for \((m_1, m_\delta)\) we have \( m_\delta \sim m_1^3 \) on average (not shown). Focusing on the \((m_1, m_\nu)\) case we can pinpoint why such nonlinear dependencies cause degeneracy. Since \( m_\nu \sim m_1^2 \), choosing the constraints arbitrarily we are \textit{independently} setting both the average of \( m_1 \) and its spread \( \sigma = (\langle m_1^2 \rangle - \langle m_1 \rangle^2)^{\frac{1}{2}} \).

This is shown most directly by looking at an ERG \((m, m_1^2)\) model (see Fig. 2). Since the network is finite, the spread \( \sigma \) can be tuned from a small value corresponding to a unimodal distribution for \( m_1 \), Fig. 3(a) and (c), to its maximum Fig. 3(d) and (f), where the probability mass is bimodal, hence causing degeneracy. Note, a linear relation between the variables will not cause degeneracy. This suggests to

\[\begin{align*}
\tilde{N}(\xi) &= N(F^{-1}(\xi)) \\
\end{align*}\]

\text{Fig. 2.} \( N(m) \) (black dots) and the domain of the values that \( m \) can take (purple shading) in a given ERG model. The orange line shows the relationship \( m_\nu \sim m_1^2 \). (a) ERG \((m_1, m_\nu)\), in \((m_1, m_\nu)\) space. Note the large shaded region where there are no realizable graphs. (b) ERG \((m_1^2, m_\nu)\), in \((m_1^2, m_\nu)\) space. The domain of the averages and the realizable graphs almost coincides. (c) ERG \((m_1^2, m_\nu)\) in \((m_1, m_\nu)\), compare with (a).
Recall that in the original (degenerate) ERG($m$) we had $\langle m \rangle = m^0$ precisely, by definition. However, the new model ERG($\xi$) is constrained by $\langle \xi \rangle_{\xi} = F(m^0) = \xi^0$, where the subscript $\xi$ indicates averages in ERG($\xi$). Here $\langle m \rangle_{\xi} \neq m^0$, yet $\langle m \rangle_{\xi} \approx m^0$ will hold. Let $k^0$ denote the Lagrange parameters in the ERG($\xi$) model. For the $i$th component, the difference is on the order of $\frac{1}{2} \sum_{\xi} (\xi - \xi^0)^T H[F^{-1}_i(\xi^0)] (\xi - \xi^0) |p(\xi; k^0)| \leq \lambda \lambda \| H[F_i^{-1}(\xi)] \| \| \xi \|_2 \| \lambda \| H[F_i^{-1}(\xi)] \| \| \xi \|_2$, where $H[F_i^{-1}(\xi)]$ is the Hessian of $F_i^{-1}(\xi)$ computed in $\xi$ and $\| \|_2$ is the spectral norm. Since ERG($\xi$) is non-degenerate, $p(\xi; k^0)$ will be concentrated around $\xi^0$, in a region small compared to $L$, and additionally, over this region the variability of $F$ is small ($F$ straightens the whole domain $\mathcal{D}$, varying significantly only over distances comparable to $L$). Thus, while this transformation leads to minor differences, it resolves the degeneracy problem and the samples are with high probability from the neighborhood of graphs for which the given constraints are typical.

Validation.—In the following we test the method on the well-known Zachary’s karate club (ZKC) experimental dataset $22$, which describes a network $G^0$ of friendships at a university karate club. It has $N = 34$, $m^0_\| = 78$, $m^0_\triangle = 528$ and $m^0_\Diamond = 45$. Using Markov Chain Monte Carlo (MCMC) sampling and a stochastic root finding method, we fitted the ERG($m_i, m_\triangle$) model to $G^0$ obtaining $\beta^0_\| = 2.610$, $\beta^0_\triangle = -0.08125$ and the degenerate distribution $p(m_i, m_\|; \beta^0_\|, \beta^0_\triangle)$ shown in Fig. 4(a). We next fitted the model ERG($\xi = m_i, \xi_\triangle = \sqrt{m_\triangle}$), obtaining $k^0_\| = 3.625$ and $k^0_\triangle = -7.998$ and a non-degenerate distribution $p(m_i, m_\|; k^0_\|, k^0_\triangle)$, shown in Fig. 4(b). The averages are summarized in Table I. Even though here we are solving for $\langle \sqrt{m_\triangle} \rangle_{\xi} = \sqrt{m^0_\triangle}$, we expect that $\langle \sqrt{m_\triangle} \rangle_{\xi} \approx m^0_\triangle$. This is confirmed in the $(m_\|)$ column of Table I. Note that due to the degeneracy of ERG($m_i, m_\triangle$), its prediction for $\langle \sqrt{m_\triangle} \rangle^2$ is $370$, far from $528$, whereas ERG($m_i, \sqrt{m_\|}$) predicts both quantities very well.

Let us now consider another measure, the number of triangles $m_\triangle$. To the extent in which $m_i^0$ and $m_\triangle^0$ determine $m_\triangle$, the corresponding ERG model should predict $m_\triangle$ as well. Unsurprisingly, ERG($m_i, m_\|)$ produces a bimodal distribution $p(m_\|, m_\triangle)$, Fig. 4(c)-(d) and predicts $(m_\triangle) = 78$, far from $45$. Additionally, $45$ and $78$ are produced with low probability in the ERG($m_i, m_\|$) model (see Fig. 4(d)). The ERG($m_i, \sqrt{m_\|}$) convexified model, however, predicts $(m_\triangle) = 40$, and both $40$ and $45$ are produced with high probability in this model (see Fig. 4(e)).

The maximum entropy method imposes average constraints, with some expectation of easing the hard and often intractable problems of the sharp constraints based approach. However, as we have shown, the fundamental problem that often hinders the applicability of the maximum entropy method is traced back to the hardest problem type of the sharp constraint approach, namely to counting type problems (density of states).

We thank L. Székely and K. Bassler for discussions. This work was supported in part by grant No. FA9550-12-1-0405 of the U.S. Air Force Office of Scientific Research, the Defense Advanced Research Projects Agency and the Defense Threat Reduction Agency Award HDTRA 1-09-1-0039.

TABLE I. Averages of measures in the fitted ERG models.

| $\langle m_i \rangle$ | $\langle m_\| \rangle$ | $\langle \sqrt{m_\|} \rangle^2$ | $\langle m_\triangle \rangle$ |
|----------------------|----------------------|----------------------|----------------------|
| $G^0$ (ZKC)          | $78 \pm 0.5$         | $530 \pm 9$          | $370 \pm 4$          |
| ERG($m_i, m_\|)$     | $77.8 \pm 0.5$       | $530 \pm 9$          | $77.7 \pm 2.3$       |
| ERG($m_i, \sqrt{m_\|}$ | $77.9 \pm 0.5$       | $530 \pm 2.7$        | $527.3 \pm 2.5$      | $39.5 \pm 0.3$ |

We thank L. Székely and K. Bassler for discussions. This work was supported in part by grant No. FA9550-12-1-0405 of the U.S. Air Force Office of Scientific Research, the Defense Advanced Research Projects Agency and the Defense Threat Reduction Agency Award HDTRA 1-09-1-0039.

The maximum entropy method imposes average constraints, with some expectation of easing the hard and often intractable problems of the sharp constraints based approach. However, as we have shown, the fundamental problem that often hinders the applicability of the maximum entropy method is traced back to the hardest problem type of the sharp constraint approach, namely to counting type problems (density of states).
[1] E. T. Jaynes, Phys. Rev. 106, 620 (1957).
[2] E. T. Jaynes, Phys. Rev. 108, 171 (1957).
[3] C. E. Shannon, Bell System Tech. J. 27, 379 (1948).
[4] Z. M. Saul and V. Filkov, Bioinformatics 23, 2604 (2007).
[5] G. Yeo and C. B. Burge, J. Comput. Biol. 11, 377 (2004).
[6] M. Ercsey-Ravasz, N. T. Markov, C. Lamy, D. C. Van Essen, K. Knoblauch, Z. Toroczkai, and H. Kennedy, Neuron 80, 184 (2013).
[7] S. J. Phillips, R. P. Anderson, and R. E. Schapire, Ecol. Model. 190, 231 (2006).
[8] J. Harte, Maximum Entropy and Ecology (Oxford University Press, 2011).
[9] P. Fronczak, A. Fronczak, and J. A. Holyst, Phys. Rev. E 75, 1 (2007).
[10] A. Wimmer and K. Lewis, Am. J. Sociol. 116, 583 (2010).
[11] F. M. Bass, J. Marketing Res. 11, 1 (1974).
[12] S. Gull and G. Daniell, Nature 272, 686 (1978).
[13] U. Skoglund, L. G. Olofstedt, R. M. Burnett, and G. Bricogne, J. Struct. Biol. 117, 173 (1996).
[14] R. Rosenfeld, Comput. Speech Lang. 10, 187 (1996).
[15] W. Holland, Paul and S. Leinhardt, J. Am. Stat. Assoc. 76, 33 (1981).
[16] D. Strauss, SIAM Review 28, 513 (1986).
[17] J. Park and M. E. J. Newman, Phys. Rev. E 70, 066146 (2004).
[18] J. Park and M. E. J. Newman, Phys. Rev. E 70, 1 (2004).
[19] J. Park and M. E. J. Newman, Phys. Rev. E 72, 026136 (2005).
[20] G. L. Robins, P. Pattison, Y. Kalish, and D. Lusher, Soc. Networks 29, 173 (2007).
[21] P. Fronczak, A. Fronczak, and M. Bujok, Phys. Rev. E 88, 032810 (2013).
[22] B. Bollobás, Random Graphs, 2nd ed. (Cambridge University Press, 2001).
[23] H. Kim, Z. Toroczkai, I. Miklós, P. Erdős, and L. Székely, J. Phys. A 42, 392001 (2009).
[24] C. Del Genio, H. Kim, T. Z., and K. Bassler, PLoS ONE 5, e10012 (2010).
[25] E. Czabarka, A. Dutle, P. Erdős, and I. Miklós, http://arxiv.org/abs/1302.3548 (2013).
[26] M. Jerrum, L. Valiant, and V. Vazirani, Theoretical Computer Science 43, 169 (1986).
[27] V. Vazirani, Approximation Algorithms (Springer, 2003).
[28] See Supplemental Material at [URL will be inserted by publisher] for animated plots of N and p.
[29] S. Boyd and L. Vandenberghe, Convex Optimization (Cambridge University Press, 2004).
[30] Equivalent to the usual definition for strict concavity $g(\lambda x_1 + (1-\lambda)x_2) < \lambda g(x_1) + (1-\lambda)g(x_1), \forall x_i \in \mathcal{D}, \lambda \in (0,1)$ with $g = \ln N$.
[31] T. A. B. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock, Sociol. Methodol. 36, 99 (2006).
[32] W. Zachary, J. Anthropol. Res. 33, 452 (1977).