Modeling Partitions of Individuals

Marion Hoffman1, Per Block2,3, and Tom A. B. Snijders4,5

Abstract
Despite the central role of self-assembled groups in animal and human societies, statistical tools to explain their composition are limited. The authors introduce a statistical framework for cross-sectional observations of groups with exclusive membership to illuminate the social and organizational mechanisms that bring people together. Drawing from stochastic models for networks and partitions, the proposed framework introduces an exponential family of distributions for partitions. The authors derive its main mathematical properties and suggest strategies to specify and estimate such models. A case study on hackathon events applies the developed framework to the study of mechanisms underlying the formation of self-assembled project teams.

Keywords
exponential families, stochastic partitions, statistical modeling, social groups, self-assembled groups

INTRODUCTION
The Study of Self-Assembled Groups
Members of gregarious species have a tendency to come together in groups; this is especially pertinent in humans. Often, the composition of these groups emerge from members’ voluntary decisions, thus crystallizing socializing preferences in social groups, or goal-oriented behaviors in the case of task-oriented groups. In some cases, group membership is exclusive: every individual can only be a member of one group. This exclusivity might result from physical and temporal constraints (e.g., when group boundaries are defined by physical gathering) or structural rules (e.g., group overlap is often forbidden when groups compete for some goal). In such cases of self-assembled and exclusive groups, the decision to group with certain individuals rather than others can depend on important social mechanisms that structure the organization of a

1Institute for Advanced Study in Toulouse, University of Toulouse 1 Capitole, Toulouse, France
2Department of Sociology, University of Zürich, Zurich, Switzerland
3Leverhulme Centre for Demographic Science, Department of Sociology, and Nuffield College, University of Oxford, Oxford, UK
4Department of Sociology, University of Groningen, Groningen, the Netherlands
5Nuffield College, University of Oxford, Oxford, UK

Corresponding Author:
Marion Hoffman, Institute for Advanced Study in Toulouse, University of Toulouse 1 Capitole, 1 Esplanade de l’Université, Office T.461, 31000 Toulouse, France
Email: marion.hoffman@iast.fr
community. The present article introduces a statistical framework to model and explain observations of self-assembled exclusive groups, with a view to better understand the mechanisms underlying their formation.

Examples of self-assembled exclusive groups are numerous, ranging from mammal herds in the wild to player squads in online games. Numerous situations require individuals to organize themselves into such groups, to execute an action or acquire a resource. In the animal kingdom, many species gather into flocks, herds, or schools for traveling purposes (Okubo 1986; Reynolds 1987) and predators assemble packs for hunting (Creel and Creel 1995; Gittleman 1989). In the human world, children groups gather in the schoolyard to engage in common activities (Moody 2001), sports teams emerge to provide opportunities for shared free-time activities (Lazarsfeld and Merton 1954; Putnam 2000), and project teams assemble spontaneously to tackle organizational tasks (Guimera et al. 2005; Zhu, Huang, and Contractor 2013). In this article, we use the example of human groups, and in particular project teams in the empirical illustration.

The existence and composition of social groups have a crucial role in determining societal outcomes. Seminal sociological works recognize that by coming together in groups, individuals influence each other’s cognition, affective structures, and individual outcomes (Homans 1950; Lewin, Heider, and Heider 1936; Parsons 1949). Various theories develop concepts for group settings, such as social circles (Simmel and Hughes 1949), social foci (Feld 1982), social settings (Pattison and Robins 2002), or social situations (Block 2018). Groups lay ground for the development of social ties (Fischer 1982; Lazarsfeld and Merton 1954; Moody 2001; Simmel and Hughes 1949) and provide the context for exchange relations (Granovetter 1985), where the ability of one group member to acquire resources and social support will depend on what other members can provide. Additionally, the set of attributes present in a group, and the relations between its members, might affect some essential group outcomes. At a broader level, the formation of groups in a community can indicate and affect how different parts of the community relate to each other and segregate (Allport, Clark, and Pettigrew 1954). Investigating group formation is all the more important in instances where such outcomes are crucial to the functioning of individuals and communities. The study of these interdependent group processes calls for the development of mathematical tools tailored for this level of social unit, as argued by Lindenberg (1997).

The main aim of the model we propose is to uncover which mechanisms guide the composition of self-assembled groups in a given setting, and to assess their relative importance. Such mechanisms can fall in the categories of biological imperatives, social preferences, and exogenous constraints. Adding to the variety of their origin, the mechanisms underlying group formation can also be situated at different levels:

1. For any group member, the characteristics of the other members reflect their individual attraction toward others exhibiting some particular attribute.
2. Group composition can also reflect dyadic preferences, such as the preference of individuals connected through a relationship (e.g., kinship or friendship) to belong to the same groups.
Finally, group-level mechanisms, such as the optimization of a certain combination of attributes, can guide group formation.

In the example of project teams, individuals might seek (1) teams with individuals who have similarities that promote mutual understanding and common expectations, (2) colleagues with whom they have already collaborated, or (3) teams with an efficient distribution of competencies (Skvoretz and Bailey 2016). On top of these formation mechanisms, some contexts might constrain group compositions or sizes; for example, a maximal group size might be imposed. The proposed model aims to shed light on the role of these diverse factors in group formation processes while taking such constraints into account.

Previous Approaches

Network Approaches. A common approach to represent group membership is to define a two-mode, or bipartite, network in which nodes on one level (i.e., individuals) are connected to a second level of nodes (i.e., groups). We review here the use and limitations of models for such representations.

First, permutation test techniques and models such as the quadratic assignment procedure proposed by Krackhardt (1988) can be used to investigate whether some combinations of attributes within groups are more likely than others in a bipartite network. However, these models condition on group structure (i.e., the distribution of group sizes) and cannot be used to investigate the factors explaining the distribution of group sizes. Moreover, interpreting the effect of covariates conditional on group structure can be problematic when these covariates are potentially responsible for the group structure itself (e.g., if an attribute explains both the size of groups and homophily within the groups). In a similar vein, Ruef (2002) proposed a method to test combinations of categories of one attribute within individual subsets of individuals, conditional on the size of the given subset.

Alternatively, the exponential random graph model (ERGM) leverages the capabilities of exponential family models (Sundberg 2019) and uses techniques from spatial statistics (Besag 1974) and graphical modeling (Lauritzen 1996) to capture more complex dependencies between membership ties (Lusher, Koskinen, and Robins 2013; Schweinberger et al. 2020). The ERGM can be used to model both attribute and structural dependencies, such as the propensity of individuals to join groups with whose members they already share other group memberships. Theoretically, it is possible to restrict the support of an ERGM to bipartite networks with individuals’ degrees fixed to one (Morris, Handcock, and Hunter 2008), thus allowing one to model exclusivity in group membership and create some structural effects to capture group sizes. The main problem remaining is that the number and characteristics of second mode nodes should be predetermined and cannot be modeled themselves (Wang et al. 2009). One could circumvent this by setting an initial number of second mode nodes equal to the maximum number of groups, that is, the number of actors. It is then straightforward to see that different partitions would be represented by different numbers of equivalent networks. For example, for three individuals A, B, and C, the partition with three isolated
individuals would be equivalent to six bipartite networks, while the partition with A, B, and C in the same group would only be equivalent to three networks. Consequently, interpreting the structural parameters of such a constrained ERGM would be problematic. All in all, using constrained ERGMs would bring few insights to the mechanisms underlying the number and size of the self-emergent groups.

Butts’s (2007) general location system (GLS) model defines a variation of the network logic that integrates the constraint of exclusive group membership. This model is tailored to observations where individuals can be assigned to only one group (or location) at a time, similar to how individuals set themselves into occupations or geographic residences. In the same vein as the ERGM, the GLS framework builds on the exponential family formalism, but it requires one to know the number and characteristics of the available groups in advance.

So far, only approaches designed for dynamic changes in group compositions over time could circumvent the issue of having to predefine the second mode nodes, by artificially creating and deleting the second mode nodes (Hoffman et al. 2020), but such procedures remain ill suited for cross-sectional observations.

**Partition Approaches.** One can circumvent the difficulty of not having predefined groups by representing groups as a partition of the set of individuals, with a partition being a division of the individuals into nonoverlapping groups. Popular partition distributions are the uniform Dirichlet-multinomial partitions defined for partitions with a maximum possible number of groups (Kingman 1978; McCullagh 2011) and Poisson-Dirichlet distributions (Pitman and Yor 1997). Such families of distributions still assume a predefined number of available groups, although the possible number of groups now sits between one and a maximal value. The extension of these models when the maximum value becomes infinite is known as the Ewens distribution (Ewens 1972; McCullagh 2011). The Ewens distribution was first applied to the problem of allele sampling in genetics (Ewens 1972), but its use, as well as the use of the related Dirichlet distributions, has spread into the fields of biodiversity (Hubbell 2001), Bayesian statistics (Antoniak 1974; Ferguson 1973), and many other fields of mathematics (Crane 2016). Interestingly, the Ewens specification also defines an exponential family (Crane 2016). One limitation of these models is that they cannot incorporate attribute and structural dependencies between group memberships in the same way ERGMs and GLS models do. This is connected to their main applications to sampling problems.

In this article, we incorporate insights from the network and partition modeling literature into a novel statistical framework suited for observations of self-assembled and exclusive groups. This framework represents groups of individuals as a partition of a set of individuals and builds on the literature on exponential families for networks to capture nontrivial dependencies between the groups composing the partition. The model allows the size and composition of groups to be the result of individual, relational, and group-level processes, and it offers the possibility to draw inferences on the processes driving the formation of groups in a certain context.
DEFINITIONS

Notation

Consider a set of \( n \) actors \( \mathcal{A} \). A partition \( P \) over \( \mathcal{A} \) represents a division of these actors into nonoverlapping subsets. Formally, \( P \) is a set of groups or blocks, denoted \( G \), that satisfies the following conditions:

\[
\bigcup_{G \in P} G = \mathcal{A},
\]

\[
\forall (G, G') \in P^2, G \neq G' : G \cap G' = \emptyset,
\]

\[
\forall G \in P : G \neq \emptyset.
\]

For convenience, we define the function \( g_P : [1, n] \rightarrow P \) returning the group of a given node:

\[
g_P(i) = G \mid i \in G.
\]

We can also transform the partition representation into the binary \( n \times n \) matrix \( X = [x_{i,j}]_{i,j \in \mathcal{A}} \) where \( x_{i,j} = 1 \Leftrightarrow g_P(i) = g_P(j) \).

Figure 1 illustrates different possible representations of a partition in comparison to the ones used in the case of networks.

In the following sections, we use the notation \( \#P \) for the number of groups in a partition \( P \), and \( \#G \) to define the size of a given group \( G \). We use the letter \( P \) when referring to a random partition, and \( p \) for the realization of a partition. To avoid any confusion, probabilities are written with the symbol \( \Pr \). We further use the letters \( a \) for individual covariates (e.g., gender, age) and \( Z \) for dyadic covariates (e.g., friendship ties).

Definition of the Partition Set

The power set of all partitions over the set \( \mathcal{A} \) is referred to as \( \mathcal{P}(\mathcal{A}) \) (or \( \mathcal{P} \) when the node set is not ambiguous). The size of \( \mathcal{P} \) is given by the Bell number (Bell 1934; Pitman 1997) and can be calculated iteratively.

In certain contexts, some partitions of the actor set might not be realistic or allowed, in which case one might only consider a subset \( \mathcal{P}' \) of the whole partition space \( \mathcal{P} \). More particularly, we consider here subsets \( \mathcal{P}' \) that only contain groups of sizes higher or equal to the minimal value \( \sigma_{\text{min}} \) and lower or equal to a value \( \sigma_{\text{max}} \). The number of partitions belonging to such subsets can be calculated similarly to the general case (for details, see Appendix A).

Relations between Partitions

For the purpose of parameter estimation and interpretation, we define three symmetric binary relations between the elements of \( \mathcal{P} \), called the \textit{merge/split}, \textit{permute}, and \textit{transfer} relations (see an illustration of these relations in Figure 2).
The merge/split relation $R_{\text{merge}}$ is the set of all unordered pairs of partitions for which one partition of the pair is obtained by merging two distinct groups in the other partition. Because these are unordered pairs, in the reverse direction this definition includes splitting one group in one partition into two groups in the other. Formally, we define $P_{-G, G'} = P \setminus \{G, G'\}$ the partition $P$ with two groups $G$ and $G'$ removed. The relation can be written as

$$R_{\text{merge}} = \{ \{P, P'\} \subseteq \mathcal{P} \mid \exists G, G' \in P : P' = P_{-G, G'} \cup \{G \cup G'\} \}.$$  

The permute relation $R_{\text{permute}}$ links partitions in which two nodes in two different groups are exchanged, while the other nodes grouping remains the same. For $i$ and $i'$, two nodes belonging, respectively, to two distinct groups $G$ and $G'$, we denote $G_{i \rightarrow i'}$. 

**Figure 1.** Possible representations of a partition over the node set $\{1, 2, 3, 4, 5, 6\}$.

**Figure 2.** Illustration of the merge/split, permute, and transfer relations for the full set of partitions over three nodes. Note: All relations are binary and symmetric.
and $G'_f\rightarrow_i$ the groups in which the nodes $i$ and $i'$ have been exchanged. Under the same notation, the relation defines the following unordered pairs:

$$\mathcal{R}^{\text{permute}} = \{\{P, P'\} \subseteq \mathcal{P} \mid \exists G, G' \in P, i \in G, i' \in G' : P' = P_G \cup \{G_{i \leftrightarrow i'}, G'_{i \leftrightarrow i'}\}\}.$$  

Finally, the transfer relation $\mathcal{R}^{\text{transfer}}$ contains the unordered pairs of partitions $\{P, P'\}$ for which partition $P$ and $P'$ are identical, with the exception of one node that belongs to a different group in $P$ and $P'$ (we can say that this node is being transferred from one group to another). Importantly, this node may be an isolate in one of the two partitions. Similarly, for a node $i$ belonging to the original node set $\mathcal{A}$, we denote $P_{-i}$ the projection of the partition on the set $\mathcal{A}\setminus\{i\}$. The relation is then defined as follows:

$$\mathcal{R}^{\text{transfer}} = \{\{P, P'\} \subseteq \mathcal{P} \mid \exists i \in \mathcal{A} : P'_{-i} = P_{-i} \text{ and } P' \neq P\}.$$  

**Definition of the Probability Distribution**

Our aim is to define a parametric set of probability distributions over $\mathcal{P}$ for a given set of actors. The parameters of this distribution should be associated with statistics relevant for the hypotheses under consideration on the processes resulting in the observed partition. As outlined earlier, such hypotheses can be associated with the structure of the partition (i.e., the number of groups and their sizes) or the distribution of actors’ attributes within the groups.

The class of exponential distributions allows such parametrization in a straightforward way (Sundberg 2019). We propose an exponential family with support $\mathcal{P}$ (or a subset $\mathcal{P}^0$). This family is defined for an identity base measure, a vector of natural sufficient statistics $s(P) = (s_k(P))_{k \in K}$, and a canonical parameter vector $\alpha = (\alpha_k)_{k \in K}$. It is expressed by

$$\Pr_\alpha(P = p) = \frac{\exp\left(\sum_k \alpha_k s_k(p)\right)}{\kappa_\mathcal{P}(\alpha)}.$$  

where the normalizing constant $\kappa_\mathcal{P}(\alpha)$ is defined by

$$\kappa_\mathcal{P}(\alpha) = \sum_{P \in \mathcal{P}} \exp\left(\sum_k \alpha_k s_k(P)\right).$$  

This formulation mirrors the definition of an ERGM when considering a partition instead of a graph distribution (for more details on ERGMs, see Lusher et al. 2013; Robins, Pattison, et al. 2007).

Some special cases of this exponential family are related to well-known distributions. Naturally, the model defined without any sufficient statistic generates the uniform distribution over the partition set $\mathcal{P}$. The Ewens distribution (Ewens 1972; McCullagh 2011) is defined for a positive parameter $\lambda$ as follows:

$$\Pr_\lambda(P = p) = \frac{\Gamma(n - 1) \lambda^\#p \prod_{G} (\#G - 1)!}{\Gamma(n + \lambda - 1)}.$$  

Hoffman et al.
with $\Gamma$ being the gamma function. As shown in Appendix B, this definition is equivalent to the following formulation of equation (1) with the parameter vector $\alpha = (\log(\lambda), 1)$:

$$\Pr_\alpha(P = p) = \frac{\exp(\alpha_1 \#p + \alpha_2 \sum_{G \in \mathcal{G}} \log((\#G-1)))}{\kappa_p(\alpha)}.$$  \hspace{1cm} (4)

### MODEL SPECIFICATION

#### Sufficient Statistics

Graphical modeling with dependence graphs is a useful technique for specifying exponential family distributions (Lauritzen 1996). In the network literature, this technique was introduced for Markov graphs by Frank and Strauss (1986) and later developed for ERGMs (Robins and Pattison 2012; Wasserman and Pattison 1996). Dependence graphs then capture the dependence structure of the tie variables, and this structure can inform the choice of relevant sufficient statistics, by virtue of the Hammersley-Clifford theorem (Hammersley and Clifford 1971; see also Besag 1974). However, graphical modeling is ill suited to study partition models, because the dependence graph of group variables with the nonoverlapping constraint is not straightforward. Instead, we take inspiration from the statistics and the independence assumptions used in other related statistical models, in particular partition models (i.e., Ewens and Dirichlet partitions) or Dirichlet models. Extending the statistics used in the Ewens formula, we show that statistics defined as sums of group attributes can model a wide range of partition properties. The independence properties of count statistics are described in the section “Independence Properties of the Distribution.”

#### Structural Statistics

Structural statistics aim to accurately model the observed group sizes and their dispersion in a given partition. To understand which statistics can be used, we calculate the expected distribution of group sizes in random partitions of 10 nodes for different statistics related to group sizes. Having $n = 10$ allows us to enumerate the number of partitions with specific group statistics and directly calculate their probabilities (for more details on these probabilities, see the section “Computation of the Normalizing Constant of the Distribution”). Increasing the number of nodes does not affect the behavior of the presented statistics.

The first relevant statistic to model group sizes is the number of groups (i.e., the cardinality of the partition; see Figure 3a) $s_1(P) = \#P$, as it is used in the Ewens formula (see equation 4). Figure 4a shows that low values of $\alpha_1$ favor partitions with large groups of 10, 9, or 8 nodes, while high values favor many small groups of 1, 2, or 3 nodes. Figure 4b shows that as $\alpha_1$ increases, the expected number of groups increases, and so does the expected number of singleton groups. The expected number of groups of size 10, that is, trivial one-group partitions, decreases, and the expected prevalence of the intermediate group sizes 2 to 9 is unimodal, and assume their maxima for values of $\alpha_1$ that decrease with group size. Figure 4c further shows that the probability for a random node to belong to large groups decreases when $\alpha_1$ increases. Finally, Figure 4d shows that the distribution of group sizes stochastically decreases...
Figure 3. Illustration of the calculation of introduced statistics on the basis of counts for a given partition $p = \{\{1, 2, 5\}, \{3, 4\}, \{6\}, \{7, 8, 9, 10\}\}$ with a binary covariate (actor’s shape). Note: The black dashed elements are counted to get the statistics value for (a) number of groups $s_1(p) = 4$; (b) squared group sizes, that is, each unordered dyad must be counted twice plus the number of nodes $s_3(p) = 30$; (c) number of dyads within groups that are identical on shape $s_{dyadic\ homophily}(p) = 4$; and (d) number of ordered dyads (e.g., directed ties) within groups that include one square $s_{dyadic\ sociability}(p) = 8$.

Figure 4. Distribution of group sizes in a random partition defined by a model with 10 nodes and one sufficient statistic $s_1(P) = \#P$, as a function of the parameter $\alpha_1$.
Note: (a) Expected group size of a given node, (b) expected number of groups of a given size, (c) probability function for the size of a given node for three values of $\alpha_1$, and (d) expected distribution of group sizes for three values of $\alpha_1$. 
We conclude that the number of groups is a simple and efficient way to model the central tendency of group sizes in a partition.

Another important feature of the group size distribution is its dispersion or skewness. We first use the statistic in definition (4) of the Ewens distribution:

\[ s_2(P) = \sum_{G \in P} \log \left( \frac{\#G}{C_0} \right) ! \]

Because the Ewens model can reproduce a “richer-get-richer” effect on group sizes, we can expect it is the result of this term being included in the model (to understand this effect, one can use the link outlined by McCullagh 2011 and Crane 2016 between the Ewens formula and a Chinese restaurant process in which individuals join tables with a probability proportional to the number of individuals at that table). We calculate the size distribution for partitions over 10 nodes for a model containing the two statistics \( s_1 \) and \( s_2 \) by varying the parameter \( \alpha_2 \). To fix the first statistic, we determine the value \( \alpha_1 \) that maintains the expected value of \( s_1 \) equal to 4 for each predetermined \( \alpha_2 \). This means we explore the distribution of expected group sizes for a constant expected number of groups. Figure 5a shows the expected distribution. The dispersion of sizes increases with the parameter value for the statistic \( s_2 \).

Another intuitive statistic for modeling the skewness of the size distribution is the sum of squared sizes:

\[ s_3(P) = \sum_{G \in P} \#G^2 \]

It is equal to the sum of the elements of the matrix representation \( X \) of the partition (see Figure 3b). The group size distributions obtained for this statistic are shown in

---

**Figure 5.** Distribution of group sizes for a random partition defined by a model with 10 nodes and two sufficient statistics for three values of the second parameter (see the text for the determination of the first parameter): (a) \( s_1(P) = \#P \), \( s_2(P) = \sum_{G \in P} \log (\left( \text{#}G - 1 \right)) ! \); and (b) \( s_1(P) = \#P \), \( s_3(P) = \sum_{G \in P} \#G^2 \).
Figure 5b. Once again, increasing the value $\alpha_3$ can increase the dispersion of sizes in the random partition. Choosing between $s_2$ and $s_3$ to model size dispersion is then a practical matter of which one more accurately represents the structure of the observed partition. In case the distribution of group sizes cannot be approximately reproduced by the above parameters, or if a particular group size might be over- or underrepresented for exogenous reasons, the number of groups of particular sizes can be added as a sufficient statistic.

At this point, it is important to mention that some estimation issues coined as *degeneracy* or *near degeneracy* in the ERGM literature (Handcock 2003; Lusher et al. 2013; Robins, Snijders, et al. 2007; Snijders et al. 2006) might ensue from the use of certain statistics combinations in this model. This is the case for the previous models defined for $S=(s_1, s_2)$ and $S=(s_1, s_3)$. As a result, some estimated models will correspond to extreme distributions that concentrate their probability mass on a few extreme partitions, such as the model with only one group and the model only containing singletons rather than accurately reflecting the observed statistics. In some cases, this behavior might be interesting in itself, if we expect that a sharp phase transition between two extremes (i.e., between one situation where individuals are completely isolated and one where everyone is in the same group) is theoretically reasonable. In other cases, a degenerate model will point to some misspecification, and it might prove useful to have a different operationalization of size dispersion. For example, one might use a weighted sum over all group sizes of the number of cliques of a given size. Weights could be defined as decreasing in a similar way as the “geometrically weighted edgewise shared partners” effect proposed by Snijders et al. (2006) and Hunter (2007). Most observations on degeneracy made in the case of ERGMs can be extended to the model presented here.

Finally, the question of how these statistics and their parameters behave for different numbers of nodes remains unknown. Further work could address this issue, which is also encountered for ERGMs (see Butts 2022; Butts and Almquist 2015; Krivitsky, Handcock, and Morris 2011).

**Statistics for Covariates.** The influence of individual covariates on the formation of relational ties has been widely investigated in social networks, starting with the fundamental idea of *homophily* (McPherson, Smith-Lovin, and Cook 2001; Rivera, Soderstrom, and Uzzi 2010) stating that similar individuals are more likely to be connected. In the case of a dyad, homophily can be operationalized as a dyadic variable indicating whether the actors have the same (or similar) attributes. Including this mechanism in the current model requires us to extend the concept of homophily to the group level.

The most direct extension of dyadic homophily is to consider a group as homophilous if the similarity within all dyads of actors within the group is high. For example, hackathon participants might be more likely to form teams in which everyone is about the same age. If we define a dyadic similarity index $\text{sim}_{i,j}$ for two actors $i$ and $j$, we can operationalize this type of homophily by counting the sum of similarity indexes among dyads in all groups:

$$s_{\text{dyadic homophily}}(P) = \sum_{G \in P} \sum_{i,j \in G} \text{sim}_{i,j}.$$
For a binary attribute, the similarity index can take the value of 1 when two actors have the same attribute, and 0 otherwise. This amounts to counting the number of ties between identical individuals in the network representation of the partition, as illustrated in Figure 3c. For a categorical and ordered attribute, or a continuous attribute, we can simply use the absolute difference between actors’ attributes (in that case, a positive parameter for this effect shows heterophily, and a negative one homophily).

Alternatively, homophily can operate at the group level. That would mean that the similarity of dyads does not matter as much as the distribution of the attributes within groups. For example, hackathon participants might form teams that are strictly non-mixed, in terms of language spoken or gender. In that case, a similarity index \( \text{sim}_G \) is defined for any group \( G \), and the relevant statistic becomes

\[
\text{s}_{\text{group homophily}}(P) = \sum_{G \in P} \text{sim}_G.
\]

For a binary attribute, the similarity index of a group can be a simple indicator that equals 1 when all group members have the same attribute. One can also use the range of the attribute values in the group in the case of a continuous attribute (as an indication of the concentration of values in a certain interval) or the number of different values of the attribute in the group in the case of a categorical variable. Alternatively, the within-group variance of the attribute (e.g., the product \( p(1 - p) \), when \( p \) is the proportion of a binary attribute) can also be a good choice for \( \text{sim}_G \).

Many variations on the idea of similarity (or dissimilarity) within group members can be constructed. In the dyadic definition, the count of similar ties could be replaced by the count of individuals who have at least a certain number of similar individuals in their groups, if similarity between group members only matters until a certain threshold. In the group definition, one can consider a statistic counting groups with a certain combination of attributes, if complementarity is thought to be more important than similarity. Furthermore, various homogeneity indexes, such as the Blau’s heterogeneity index (Blau 1977), could be used.

In practice, these definitions might lead to a bias toward certain numbers of groups (in the case of \( \text{s}_{\text{group homophily}} \)) or sizes of groups (in the case of \( \text{s}_{\text{dyadic homophily}} \)). It can be important either to control for the number of groups and ties (with the structural statistics \( s_1 \) and \( s_3 \)) or to normalize these statistics by the number of dyads in each group (e.g., for \( \text{s}_{\text{dyadic homophily}} \)) or the size of the group (e.g., for \( \text{s}_{\text{group homophily}} \)) to increase comparability, especially in empirical cases of heterogenous group sizes. Finally, the effect of a dyadic covariate \( Z \) (e.g., friendship) can simply be added as

\[
\text{s}_{\text{dyadic covariate}}(P) = \sum_{G \in P} \sum_{i,j \in G} Z_{i,j}.
\]

**Mixed Statistics.** Mechanisms related to both covariates and structural features can be included in the current framework. This includes a translation of the network concepts of *sociability* or *aspiration* (Snijders and Lomi 2019), defined as actors’ tendencies with a high attribute to send or receive more ties, respectively. For groups, these mechanisms translate into the propensity of actors scoring high on an attribute vector \( a \) to be in larger groups. For example, extraverted individuals might be more likely to
be found in larger groups, which can be modeled by counting intragroup ties to these individuals with the following sociability statistic:

\[ s_{\text{dyadic sociability}}(P) = \sum_{G \in P} \sum_{(i, j) \in G} a_{ij}. \]

This is illustrated by Figure 3d in the case of a binary attribute. Alternatively, the sum of individual attributes can be replaced by the highest value of the attributes within groups to switch to a group-level definition:

\[ s_{\text{group sociability}}(P) = \sum_{G \in P} \#G \max\{a_i\}_{i \in G}. \]

**Link to Statistics in Other Models.** Given that the proposed model builds on the same methodological foundations as the bipartite ERGM and the GLS, it is not surprising that some of the statistics defined above have analogous formulations in these two models. For example, the sum of square group sizes (\(s_3\) in the section “Structural Statistics”) is equivalent to the in-star parameter in bipartite ERGMs, which is defined as \(\frac{1}{2} \#G(\#G - 1)\), using our notation. Similarly, the statistics defined in the section “Statistics for Covariates,” can be reformulated as a case of “object homogeneity/heterogeneity potential” for the GLS (Butts 2007, section 3.2.2). An equivalent formulation for the bipartite ERGM is discussed by Bomiriya (2014) and, with some modifications, applied by, for example, Balest et al. (2019).

Note that we do not delve into the theoretical implications of using the statistics proposed in the prior two sections. Further work should refine the formulation of these statistics, potentially considering how qualitatively similar questions have been or can be addressed in bipartite ERGMs and GLS models.

**Independence Properties of the Distribution**

Classical methods of graphical modeling (Lauritzen 1996) are ill suited for representing dependence assumptions in partition models. However, we can use other concepts to discuss independence properties of the model.

Kingman (1978) established the property of *consistency* for the Ewens sampling formula. This concept, translated to our case, represents that if a random partition of a given actor set \(\mathcal{A}\) has a Ewens distribution with parameter \(\lambda\), the partition restricted to any subset of \(\mathcal{A}\) will also have a Ewens distribution with the same parameter. As shown by simple counter-examples in Appendix C, most models defined for the statistics presented above fail to fulfill this condition. The Ewens formula is a special case in that regard, as consistency is a critical property for the study of population samples, yet much less so in the case of complete observations in our case.

A second relevant concept is *neutrality*, as introduced by Connor and Mosimann (1969) to study distributions of proportions of a fixed quantity. Such variables are defined as a strictly positive vector \((X_1, X_2, \ldots, X_n)\) with \(X_1 + X_2 + \ldots + X_n = q\) where \(q\) is constant. Each variable will never be independent from the others, as it can be expressed as a linear combination of the others. To remedy this, Connor and Mosimann (1969) introduce the concept of neutrality that defines the proportion \(X_1\),
for example, as neutral if it is independent of the vector \((X_2/(q - X_1), \ldots, X_n/(q - X_1))\). This property allows us to ignore one or several proportions to study the others. For example, prior work shows that neutrality of all proportions characterizes the Dirichlet distribution (Connor and Mosimann 1969; Geiger and Heckerman 1997).

Although the concept of neutrality was initially defined for proportion vectors, its extension to partitions can help us understand how the composition of a subset of the partition might affect the rest of the partition. Let \(P\) be a random partition over a set \(\mathcal{A}\), and \(\mathcal{A}'\) a subset of \(\mathcal{A}\), with complement set \(\mathcal{A}^c\). We further define \(\pi\) and \(\pi^c\) as the respective projections of partitions in \(\mathcal{P}(\mathcal{A})\) over \(\mathcal{A}'\) and \(\mathcal{A}^c\).

We define a distribution to be neutral if and only if the projections of \(P\) on \(\mathcal{A}'\) and \(\mathcal{A}^c\) are independent under the condition that any group of \(P\) is either in \(\mathcal{A}'\) or \(\mathcal{A}^c\). This condition is equivalent to having \(P\) as the union of its two projections: \(P = \pi(P) \cup \pi^c(P)\). A distribution is neutral if and only if

\[
\Pr_n(P = \pi(P) \cup \pi^c(P)) = 
\Pr_n(\pi(P) = \pi(p)) \times \Pr_n(\pi^c(P) = \pi^c(p)).
\]

We show in Appendix C that this property holds for any model specified for statistics \(s_k\) defined as sums of real functions of the groups of \(P\). Notably, all statistics proposed in the previous section and used in our analyses later are of this form, which allows us to interpret their associated parameters using simple log-odds ratios using the relations shown in the section “Relations between Partitions” (see “Results”).

**COMPUTATION OF THE NORMALIZING CONSTANT OF THE DISTRIBUTION**

As shown in the rewiring of the Ewens formula in equation (3), some model specifications induce a simplification of equation (1) into more tractable forms. This allows a direct evaluation of equation (1) in these cases, which can be leveraged to approximate the likelihood of more complex specifications.

For a model specified only with the statistic \(s_1(P) = \#P\) for a set of \(n\) nodes, we can use the Stirling numbers of the second kind (Riordan 1958) to derive a simple formulation of the normalizing constant \(\kappa_{\mathcal{P}}\). The Stirling number \(\{\begin{array}{c} n \\ m \end{array}\}\) is the number of partitions with \(m\) groups, in other words, the number of partitions for which \(s_1(P) = m\) (Pitman 1997). We can therefore sum over all possible values \(m\) and get the direct expression:

\[
\kappa_{\mathcal{P}}(\alpha_1) = \sum_{m=1}^{n} \{\begin{array}{c} n \\ m \end{array}\} \exp(\alpha_1 m).
\]

More interestingly, we can calculate the normalizing constant of any model containing statistics of the form

\[
s_k(P) = \sum_{G \in P} f_k(\#G),
\]
where \( f_k \) are functions of the block sizes. For such models, the sufficient statistics define an exchangeable distribution (McCullagh 2011) that does not depend on the labeling of the nodes. We define \( \kappa_n \) as the normalizing constant of these models for a given \( \alpha \) on any set of \( n \) nodes. This constant can be constructed as a recursive sequence and computed with the following formulas:

\[
\begin{align*}
\kappa_0 &= 1, \\
\kappa_1 &= \exp \left( \sum_{k \in K} \alpha_k f_k(1) \right), \\
\kappa_{n+1} &= \sum_{i=0}^{n} \binom{n}{i} \exp \left( \sum_{k \in K} \alpha_k f_k(n+1-i) \right) \kappa_i.
\end{align*}
\]

As mentioned in the section “Relations between Partitions,” some analyses might require restricting the outcome space to subsets \( \mathcal{P}' \) containing partitions with groups of sizes between \( \sigma_{\text{min}} \) and \( \sigma_{\text{max}} \). Equation (6) can also be used in this case by replacing the Stirling numbers \( \left\{ \frac{n}{m} \right\} \) by an extension defined by the number of partitions in \( m \) blocks with all block sizes belonging to \([\sigma_{\text{min}}, \sigma_{\text{max}}] \) (Appendix A describes a recursive formula for these numbers). More generally, equation (8) can also be extended to the case of size restrictions. The proof for the general and restricted case can be found in Appendix D.

**ESTIMATION**

In the case of exponential families, the maximum likelihood estimation method is equivalent to the method of moments that consists in finding the parameters under which the expected statistics of the modeled partition are equal to the observed statistics (Sundberg 2019). Such estimations, however, require the calculation of either the likelihood function or the expected statistics under the model. When the normalizing constant of a model, and therefore its likelihood, can be calculated as shown in the prior section, any optimization method, such as a Newton-Raphson method (Deuflhard 2011), can be applied to approximate the parameter value for which this likelihood is maximum. This maximum either exists and is unique, or is infinite, by virtue of the properties of convexity in exponential families (Wedderburn 1976).

As soon as a model includes statistics related to actors’ attributes, such simplifications of the normalizing constant \( \kappa_{\mathcal{P}} \) are unlikely to be found. As \( \kappa_{\mathcal{P}} \) contains a great number of terms (as given by the Bell number, see the section “Definition of the Partition Set”), the calculation of this likelihood is practically intractable for a large number of nodes. This problem can be circumvented with Monte Carlo Markov chain (MCMC) techniques, drawing inspiration from algorithms originally devised for ERGMs (Hunter and Handcock 2006; Lusher et al. 2013; Snijders 2002).

**Sampling Partitions**

As the space \( \mathcal{P} \) becomes extremely large for high values of \( n \), we can only sample the space \( \mathcal{P} \) and obtain a subset of random partitions to approximate the distribution of
partitions under a given model. MCMC methods can assist in constructing such a subset by sampling partitions from a Markov chain whose stationary distribution is the model distribution given by equation (1). A suitable algorithm for this purpose is the Metropolis-Hastings algorithm (Chib and Greenberg 1995; Hastings 1970; Metropolis et al. 1953). The Metropolis-Hastings approach consists in defining a Markov chain with transition probabilities \( q \) defined as the following product:

\[
q(p' | p) = \tilde{q}(p' | p)A(p', p).
\]

At each step in the chain, a new partition \( p' \) is proposed with probability \( \tilde{q}(p' | p) \), and it is accepted according to the acceptance ratio \( A(p', p) \).

To define the proposal distribution \( \tilde{q} \), we use a symmetric relation \( R \) on the space \( \mathcal{P} \). This relation can be one of the previously defined relations \( R_{\text{merge}}, R_{\text{permute}}, \) and \( R_{\text{transfer}} \) or a combination of them. Importantly, this relation should connect the entire outcome space; therefore, \( R_{\text{permute}} \) should not be used without at least one of the other two, because it maintains the size structure. For our analyses, we use \( R_{\text{merge}} \) for purely structural models, and a combination of relations when covariate effects are included. To know how often each relation should be used, one can try different combinations and pick the one that leads to a good mixing of the chain.

For a given partition \( p \), we propose to only move to a partition \( p' \) such that \( p \) and \( p' \) are linked by a given relation \( R \), with a uniform probability:

\[
\tilde{q}(p' | p) = \frac{1}{\# \{ p' : (p, p') \in R \}}.
\]

By using the detailed balance equation \( q(p' | p)Pr(p) = q(p | p')Pr(p') \) that ensures the convergence of the Markov chain to the desired distribution (Chib and Greenberg 1995; Hastings 1970; Metropolis et al. 1953), we get

\[
A(p', p) = \min \left( 1, \frac{Pr(p') \# \{ (p, p') \in R \}}{Pr(p) \# \{ (p', p) \in R \}} \right).
\]

As evident from Figure 2, the proposal distribution defined for relations such as \( R_{\text{merge}} \) is not symmetric. In other words, for some pairs \( (p, p') \in R \) it is the case that \( \tilde{q}(p' | p) \neq \tilde{q}(p | p') \). Therefore, it is necessary to calculate the proposal probabilities at each step of the chain to find the acceptance ratio. Deciding which relation to use depends on how fast these calculations can be made and how efficiently the algorithm covers the sampled space. Moreover, the proposal distribution has to be adapted when the set of allowed partitions is restricted, to make sure every sampled partition fits these restrictions. In certain cases, one might want to design a chain that covers a larger space and only retain correct partitions.

**Estimation Procedure**

The estimation procedure used in this study implements the Robbins-Monro algorithm (Robbins and Monro 1951) in a similar way as used in Snijders (2001, 2002) for the estimation of network models. The Robbins-Monro algorithm is a variant of the
Newton-Raphson optimization algorithm for objective functions obtained via Monte Carlo methods. Prior work shows it is a useful tool for a large range of stochastic approximation problems (Lai 2003), in particular for the maximum likelihood estimation of models that can only be analyzed by simulations (Cappé, Moulines, and Rydén 2005; Gu and Kong 1998; Gu and Zhu 2001). Note that various other algorithms have been designed for similar problems, including the Geyer-Thompson algorithm (Geyer and Thompson 1992) and the stepping algorithm by Hummel, Hunter, and Handcock (2012).

Here, the aim of the Robbins-Monro algorithm is to solve the moment equation:

\[
E_\alpha[s] = s_{obs},
\]

where \(E_\alpha[s]\) is the expected vector of sufficient statistics for the model with parameter \(\alpha\) and \(s_{obs} = s(p_{obs})\) is the vector of statistics in the observed partition \(p_{obs}\). The original \(N\)th iteration step of the algorithm consists in drawing a variable \(s_N\) from the distribution of the statistics for the model with parameter \(\alpha_N\) and updating the model parameter to

\[
\alpha_{N+1} = \alpha_N - a_N D_N^{-1}(s_N - s_{obs}).
\]

In this equation, \((a_N)\) is called the gain sequence and controls the magnitude of the optimization steps and \(D_N\) is the scaling matrix. A classic choice for the gain is \(a_N = 1/N\) and for \(D_N\) the derivative matrix \(\partial E_\alpha[s]/\partial \alpha_N\).

Using the arguments developed by Snijders (2001), our algorithm uses in place of the matrices \(D_N\) only one scaling matrix \(D_0\) calculated once for all. This scaling matrix is the covariance matrix of a sample of the model parametrized by some starting parameters, and it represents an estimation of the sensitivity of the sufficient statistics to the parameters’ variations. This is based on a result from Polyak (1990) implying that an optimal rate of convergence is obtained for the tail average of the sequence \((\alpha_N)\), as long as the sequence \((a_N)\) converges at the rate \(N^{-c}\), with \(0.5 < c < 1\).

Regarding the gain sequence, we use Pflug’s (1990) idea that it is better to keep a constant value \(a_N\) as long as the sequence \(s_N\) has not crossed the observed values \(s_{obs}\) yet. The algorithm is therefore divided in \(R\) subphases within which the value \(a_r\) is kept constant while the sequence \((\alpha_{r,N})\) is updated with the adapted steps (equation 10):

\[
\alpha_{r,N+1} = \alpha_{r,N} - a_r D_0^{-1}(s(p_N) - s_{obs}),
\]

with \(p_N\) drawn from the model parametrized by \(\alpha_{r,N}\). Importantly, the lengths of the subphases must ensure the convergence of \((a_N)\) at the rate \(N^{-c}\), and the starting parameter value for the subphase \(r\) should be the average of the sequence \((\alpha_{r-1,N})\) within the previous subphase, to satisfy the convergence conditions mentioned earlier.

In practice, the algorithm is implemented in three phases. The first phase is used to estimate the matrix \(D_0\) by sampling \(M_1\) partitions \(p_1, p_2, \ldots, p_{M_1}\) from the model defined for the starting parameters \(\alpha_0\), with the Metropolis-Hastings algorithm presented earlier. A good choice for \(\alpha_0\) is often a vector containing zeros except for parameters that can be calculated with the equations shown in equation (4). We only retain
partitions after a burn-in period and with a certain thinning interval to ensure a low auto-correlation between the sampled statistics (below 0.4 is an efficient rule of thumb). A value of a few hundreds for \( M_1 \) after thinning usually sufficed. We obtain an estimation of the expected statistics and the covariance matrix:

\[
\begin{align*}
\bar{s}_{a_0} &= \frac{1}{M_1} (s(p_1) + s(p_2) + \ldots + s(p_{M_1})) \\
\text{cov}(\alpha_0) &= \frac{1}{M_1} \sum_{m=1}^{M_1} (s(p_m)s(p_m)^T) - \bar{s}_{a_0}\bar{s}_{a_0}^T
\end{align*}
\]

The scaling matrix \( D_0 \) is usually defined as \( D_0 = \text{diag}(c\text{ov}(\alpha_0)) \). To achieve a more stable algorithm, the nondiagonal elements of \( D_0 \) can also be multiplied by a constant between 0 and 1 (our examples use 0.2). Its inverse \( D_0^{-1} \) provides the new starting estimates:

\[
\alpha_0 - aD_0^{-1}(\bar{s}_{a_0} - s_{obs}).
\]

In the second phase, we implement the iterative steps of equation (11) to calculate the sequences \( (\alpha_{r,N}) \) for \( R \) subphases. At each \( N \)th iteration, only one partition \( p_N \) is drawn from the distribution with parameter \( \alpha_{r,N} \), with the Metropolis-Hastings algorithm starting at the previously drawn partition \( p_{N-1} \). Each \( r \) subphase lasts until its length is above the minimum length of the subphase and all sampled statistics have crossed the observed values. Alternatively, it stops when \( N \) is above the maximal length of the subphase. In this study, we used the values \( R = 4 \), \( a = 0.1 \), and \( a_r = a/(2^{r-1}) \), and kept the lengths of the subphases between \( 2^{4r/3} \) and \( 2^{4r/3} + 200 \) to ensure the algorithm convergence and the crossing of statistics.

Finally, Phase 3 is used to sample \( M_3 \) partitions from the final distribution to approximate the expected sufficient statistics with the sample mean \( \bar{s}_{a_f} \) and the covariance matrix of these statistics with the sample covariance matrix. We used large values of \( M_3 \), typically between 1,000 and 2,000. We assess model convergence by calculating the sample standard deviation for each statistic separately. It is considered excellent for the \( k \)th statistic when the convergence ratio \( c_k \):

\[
c_k = \frac{s_{a_f,k} - s_{obs,k}}{SD_{\alpha_f}(s(p_1), \ldots, s(p_{M_f}))}
\]

remains between \(-0.1\) and \(0.1\), with \( SD_{\alpha_f}(s(p_1), \ldots, s(p_{M_f})) \) being the sample standard deviations. This value is aligned to the one chosen for ERGM estimation (see Snijders 2002). Furthermore, we assume that parameter estimates have an approximate multivariate normal distribution, similar to ERGMs (Lusher et al. 2013). We can therefore test the significance of the model parameters from a simple Wald test considering whether the ratio between the elements of \( \alpha_f \) and their standard errors (calculated from the inverse of the sample covariance matrix in Phase 3) are smaller than \(-2\) or larger than \(2\).

Appendix E presents pseudocode for the previously described algorithm. Implemented code, documentation (including UML charts), and an example script can
be found in the supplementary materials and the R package ERPM (http://github.com/marion-hoffman/erpm). We compared the results of the Robbins-Monro algorithm to a simple Newton-Raphson estimation in the case of a simple model for which the likelihood can directly be calculated.

Model Diagnostics

The goodness-of-fit of a model can be assessed by the calculation of auxiliary statistics (i.e., not included in the sufficient statistics of the model), similar to ERGMs (Hunter, Handcock, and Hunter 2008). By sampling from the estimated model, we can test whether the obtained distribution of such auxiliary statistics corresponds to those in the observed data.

To compare different model specifications, we can calculate log-likelihoods, in a similar way to the one proposed by Hunter and Handcock (2006) for ERGMs. This calculation is done through path sampling as presented by Gelman and Meng (1998) to estimate the log-likelihood of a model for an estimated parameter $\alpha$ when its normalizing constant $\kappa(\alpha)$ is intractable. First, we calculate the log-likelihood $\ell(\alpha_0, p_{obs})$ of the model parametrized by $\alpha_0$ in which statistics are identical to the ones in the model of interest but all parameters except the one for the statistic $s_1(P) = \#P$ are set to zero. The value of $\alpha_0$ is calculated using the equations in the section “Computation of the Normalizing Constant of the Distribution.” We can then estimate the difference between the normalizing constants $\lambda(\alpha_0, \alpha) = \kappa(\alpha) - \kappa(\alpha_0)$ by sampling $M$ models with parameters $\alpha_m = \frac{m}{M} \alpha + \frac{1 - m}{M} \alpha_0$ that produce large overlaps between the sampled distributions:

$$\hat{\lambda}(\alpha, \alpha_0) = \frac{1}{M} \sum_{m=1}^{M} (\alpha - \alpha_0)^T s_{a_m}.$$  

We finally estimate the log-likelihood of our model of parameter $\alpha$ by

$$\hat{\ell}(\alpha, p_{obs}) = \ell + (\alpha - \alpha_0)^T s_{obs} - \hat{\lambda}(\alpha, \alpha_0).$$

CASE STUDY: THE COMPOSITION OF SELF-FORMED TEAMS DURING HACKATHONS

Data

Topi and Tucker (2014) defined hackathons as “problem-focused computer programming events.” Hackathons are often designed for participating teams to solve a digital problem in a short period of time. Such events provide companies, universities, or nonprofit organizations the opportunity to harness the ideas of volunteers in exchange for rewards and funding for the winning teams (Briscoe and Mulligan 2014; Lara and Lockwood 2016). Hackathons now tackle a broad range of topics, including education, marketing, and arts (Lara and Lockwood 2016).

We collected data during two editions of a hackathon at a technical university. The events welcomed 60 and 58 participants, respectively, who divided themselves into 14
teams in both cases. Via online questionnaires, the registration process gathered participants’ individual attributes and their prior acquaintances. Registered participants were invited to the venue on a Saturday at 9 a.m. and were introduced to their tasks. They were then asked to mingle and define teams until 1 p.m. Teams could include two to five individuals in the first edition and three to five people in the second. The teams collaborated until Sunday afternoon to design and implement their solution to the hackathon challenge. At the end, a jury of experts assessed the teams’ compositions and their performance.

In the first edition, 1 team of two, 1 team of three, 5 teams of four, and 7 teams of five were formed. The 14 teams in the second edition were divided into 1 team of three, 9 teams of four, and 3 teams of five. Table 1 shows descriptive statistics for the participants’ attributes. In the first edition, 22 pairs of participants reported already knowing each other, and 23 such pairs were reported in the second edition.

### Table 1. Counts of Gender, Age, Language, Degree, and Major Attributes among Participants of the First and Second Hackathon Editions

|                  | First Edition | Second Edition |
|------------------|---------------|----------------|
| Gender           | (N = 60)      | (N = 58)       |
| Male             | 49            | 55             |
| Female           | 11            | 3              |
| Age              | (N = 43)      | (N = 54)       |
| <20              | 11            | 12             |
| 20-25            | 13            | 25             |
| 25-30            | 10            | 13             |
| >30              | 9             | 4              |
| First language   | (N = 49)      | (N = 56)       |
| Swiss German     | 16            | 16             |
| German           | 10            | 10             |
| Others           | 23            | 30             |
| Major            | (N = 60)      | (N = 58)       |
| Engineering      | 14            | 34             |
| Computer science, IT | 23      | 10             |
| Physics          | 6             | 3              |
| Mathematics      | 2             | 2              |
| Chemistry        | 3             | 5              |
| Environmental sciences | 4   | 3              |
| Other            | 8             | 1              |

Self-assembled teams for short projects are ubiquitous in organizational, educational, and recreational contexts (Contractor 2013; Falk-Krzesinski et al. 2010; Guimera et al. 2005; Zhu et al. 2013). Scholars investigating the motivations for individuals to form teams in various settings generally identify four types of mechanisms, as classified by Bailey and Skvoretz (2017): homophily, competence, familiarity, and affect.

First, homophily, that is, the tendency for similar individual to collaborate, is commonly observed in dyadic collaborations and teams (Gompers, Huang, and Wang 2017; Kalleberg et al. 1996; McPherson and Smith-Lovin 1987; Ruef, Aldrich, and Carter 2003). In our context, being of a similar age or sharing the same language could support communication within teams and may therefore have contributed to the choice
of teammates. Other attributes, such as gender or personality traits, might also have been relevant in this context but were discarded because of a lack of variation or simply because of a lack of explanatory power of the models.

Second, the competence of team members is central in teams whose aim is to achieve a given task. When forming teams, it is difficult to find the right balance between optimizing the number of skills within team members and reducing overhead costs of combining different ways of thinking or working. Previous research on self-assembled teams finds evidence for complementarity of skills (Zhu et al. 2013) and skill homophily (Gómez-Zará et al. 2019). During the hackathons, organizers strongly recommended that participants form teams with as diverse skills and knowledge as possible. We thus tested whether participants were more likely to form teams with individuals who had different university majors.

Third, individuals may be more comfortable teaming up with individuals familiar to them, with whom they may have collaborated in the past, because of shared practices or values (Bailey and Skvoretz 2017; Gómez-Zará et al. 2019; Lungeanu et al. 2018). Because some participants knew each other prior to the event and registered together, we expected to find a high number of prior acquaintance ties within the teams.

Finally, interpersonal affect (or dislike) can be a strong predictor in the choice of team partners, arguably even more important than competence (Casciaro and Lobo 2008). However, our data do not contain such information, so we did not test any related mechanism.

Results

Three models are presented for each data set. The first two models include effects related to group sizes and individual attributes; these are used to explore the specification of statistics related to age and language. The third, or final, models include the additional effect of previous acquaintances. We use these models to illustrate how the modeling framework can be used rather than provide in-depth theoretical insights. All models are reported in Table 2.

The group size distribution is modeled by the number of groups and the sum of squared sizes in the first data set. In the second data set, the second statistic is excluded because of degeneracy issues and replaced by the number of groups of size four because participants were advised, although not obligated, to form groups of this size. In both cases, we limit the allowed group sizes to a minimum of two and a maximum of five in the estimation procedure. Age homophily is modeled using the sum of age ranges in the groups (except in model 1 of the first data set). Homophily for language and major uses the sum of different attributes present in each group (except in model 1 of the second data set in the case of language). Finally, the influence of previous acquaintances is captured by the count of such ties within all teams.

In models 1 and 2 for the first data set, we compare two specifications for age homophily, using the sum of absolute differences in age among teammates (i.e., dyadic homophily effect) and the sum of age ranges in each team (i.e., group homophily effect), respectively. Both parameters are negative, indicating a tendency to form
Table 2. Estimated Parameters for the Models of the Two Hackathon Editions

| Sufficient Statistic       | Model 1       | Model 2       | Model 3       |
|----------------------------|---------------|---------------|---------------|
|                            | Estimate      | Significance  | Estimate      | Significance  | Estimate      | Significance  |
|                            | Standard Error|               | Standard Error|               | Standard Error|               |
| First edition Number of groups | -4.67         | 4.98          | -4.73         | 4.85          | -4.62         | 4.73          |
| Sum of squared sizes       | .05           | .35           | -0.06         | .34           | -0.07         | .34           |
| Age differences            | -.027         | .021          |               |               |               |               |
| Age ranges                 | -.16          | *             | .08           | -10           | .08           |               |
| Number of languages        | -.10          | .51           | -0.09         | .51           | .29           | .50           |
| Number of majors           | .05           | .47           | .02           | .48           | .29           | .50           |
| Number of acquaintances    | 2.90          | ***           |               |               | 2.90          | ***           |
| Log-likelihood             | -124.2        |               | -123.7        |               | -102.1        |               |
| Second edition Number of groups | -4.00         | 2.44          | -3.85         | 2.65          | -3.74         | 2.44          |
| Number of groups size 4    | 2.02          | ***           | 2.03          | ***           | .61           | 2.02          |
| Age ranges                 | -.29          | *             | -.27          | *             | -.12          | -18           |
| Same language pairs        | .72           | ***           | .21           |               |               |               |
| Number of languages        | -.33          | .58           | -0.49         | .61           | -0.07         | .60           |
| Number of majors           |               |               |               |               |               |               |
| Number of acquaintances    |               |               |               |               |               |               |
| Log-likelihood             | -106.8        |               | -105.9        |               | -89.8         |               |

Note: Convergence ratios for each parameter are below 0.12.
groups with low age differences. However, only the parameter of model 2 is significant, and the log-likelihood of $-124.2$ in model 1 is also higher than model 2’s $-123.7$ (these log-likelihoods can directly be compared because the number of parameters is constant across models). These points suggest that the age range specification might better explain the data at hand.

We carry out a similar exploration in models 1 and 2 of the second data set, where language homophily is specified by the number of same-language ties within teams or the number of different languages in each team. This time, both parameters are strongly significant and indicate a tendency to form groups with a high number of same-language ties or a small number of different languages. However, the slightly better log-likelihood of model 2 ($-105.9$ vs. $-106.8$) indicates that the second specification performs better.

Finally, we turn to the final models (model 3). We first discuss one-by-one the direction and significance of the parameters. To interpret the size of parameter values beyond their signs, we can use the binary relations \textit{merge/split}, \textit{permute}, and \textit{transfer} introduced in the section “Relations between Partitions” to define pairs of partitions that exhibit a unit change for a given statistic, \textit{ceteris paribus}. We can then formulate log probability ratios between partitions that are related through one of those relations and attach a quantitative interpretation to exact parameter values. This interpretation is made possible because the models presented here respect the neutrality property defined in the section “Independence Properties of the Distribution.” However, the \textit{ceteris paribus} condition is not trivial to invoke, as it is not always possible to find partition changes that only affect one statistic at a time. Such log probability ratios should therefore be interpreted with caution.

In the first data set, the negative parameter for the number of groups of $-4.62$ indicates a tendency to form fewer, and therefore larger, groups. Here, we can interpret this parameter with the log probability ratio between two partitions linked by the \textit{merge/split} relation. Specifically, model 3 predicts a partition to be about $\exp(4.62) \approx 101$ times more likely compared to the same partition with one group split into two, given that all other statistics remain constant and group sizes stay in the allowed range. This applies, for example, to the comparison of having one group of four participants compared to two groups of two participants, \textit{ceteris paribus}. The negative parameter for squared sizes of $-0.07$ shows a concentration of sizes around large sizes (i.e., four and five). However, these two effects are not significant (potentially because of the small size of the data set). All effects related to individual attributes are also insignificant (including the effect of age that is explained away by the effect of previous acquaintances from model 2 to 3). Their directions, however, suggest a tendency to form groups with low age differences, diverse languages, and diverse majors. Finally, we find a positive and strongly significant effect of previous acquaintances. For this statistic, it is more useful to invoke the \textit{permute} relation to calculate log probability ratios. Its parameter indicates that a partition obtained from a permutation of two actors that would add one acquaintance tie in a group, leaving other statistics equal, is around $\exp(2.90) \approx 18$ times more likely than before permutation.
Regarding the second data set, the negative parameter for the number of groups of \(-2.04\) in model 3 shows a tendency to form fewer groups, but this effect is not significant either. The significant parameter of 2.02 for groups of size four indicates individuals’ tendency to form more groups of this size than others. The directions of the parameters for attribute-related effects suggest a tendency to form groups with low age differences and diverse majors, as in the other data set, but with a low diversity in terms of languages. This significant effect for languages can be interpreted again with a log-odds ratio: a partition reached through the transfer of an actor that would add a new language to a group is \(\exp(1.23) \approx 3.4\) times less likely than the partition before this transfer. Once again, the effect of previous acquaintances is strongly significant, with a similar magnitude as in the first data set.

All in all, individual attributes seem to have little influence on the composition of teams during the first edition. In particular, we do not find evidence for any homophily effect, and the parameter related to majors even suggests that participants tried to diversify their teams as the organizers recommended (although this effect is not significant). On the other hand, our results suggest that forming teams with similar languages was important during the second edition. For both editions, previous acquaintances seem to have been the strongest driver of team formation.

Model Fit

In this section, we further investigate the distribution of auxiliary statistics in the estimated models to assess goodness-of-fit, following similar procedures as the ones recommended for network models (Hunter, Goodreau, and Handcock 2008). These distributions are represented by the violin plots proposed by Hintze and Nelson (1998).

We first examine in Figure 6 the distribution of group sizes in the final models (models 3). We see in Figure 6a that the distribution of group sizes is well recovered for the first data set, with the observed counts falling within the confidence intervals of the simulated models with a slight overestimation of groups of three and a slight underestimation of groups of four. Sizes are also well recovered in the second data set (6b), for which the number of groups of size four is perfectly estimated, as it was a sufficient statistic of the model.

To better understand the specification of age homophily for the first data set, we examine in Figure 7 statistics related to the distribution of ages within the teams and compare their observed values to the values simulated by models 1 and 2. First, Figure 7a shows that both proposed specifications recover equally well the distribution of age difference within ties. We assess the homogeneity of groups in terms of age by calculating the intraclass correlation coefficient of ages within groups (Figure 7b). Again, both models provide a very satisfactory fit. Finally, examination of the correlation between individuals’ ages and the size of their teams (Figure 7c) helps assess how well the models reproduce the tendency of certain ages to be present in larger groups, which is not an effect included in the model. This shows that model 2 provides a slightly better fit than model 1, which is in line with the better log-likelihood previously mentioned. This observation could be interpreted as the result of the first specification
being dependent on group sizes whereas the second is not, but this interpretation might only be valid for the data at hand.

Moving to the specification of language homophily, Figure 8 presents a similar type of auxiliary statistics, this time adapted to the categorical nature of the language attribute, for models 1 and 2 of the second data set. Figure 8a first shows that the number of ties with certain language combinations are well reproduced by the models, with the exception of ties between Swiss German speakers and other language speakers. We assess the homogeneity of groups by the average Blau’s heterogeneity index (Blau 1977) in the groups (Figure 7b). The fit of this statistic is very satisfying for both models. Finally, to understand the link between language and group sizes, Figure 8c shows the average group size of a native German speaker. Again, model 2 provides a better fit than model 1, which could explain the better log-likelihood of the model.

CONCLUSION

This article introduces the statistical framework of exponential partition models and presents its main mathematical properties. Building on the rich literature on exponential families of distributions, stochastic networks, and stochastic partitions, we show that this model can be used to uncover regularities in observations of self-assembled exclusive groups while taking into account structural dependencies between these observations. Exponential partition models can be applied to various contexts in which individuals sort themselves into groups on the basis of social preferences, opportunities, and exogenous constraints. We propose specifications to investigate a variety of mechanisms that can be situated at an individual, relational, and group level. We provide an example study case, the self-formation of hackathon teams, to illustrate some of the capabilities of the model. All code and documentation for further use of this framework can be found at http://github.com/marion-hoffman/erpm. Data for replication are available on request.

This work bridges two branches of the statistics literature, one representing systems as networks and the other as partitions. On the one hand, we augment network methods by introducing the possibility of modeling social mechanisms at the level of groups.
rather than dyads. By rethinking the mathematical representation of groups, the proposed framework allows researchers to investigate group formation as coordination processes between individuals, rather than an aggregation of dyadic ties to group entities. On the other hand, we contribute to the stochastic partition modeling literature by extending the use of such models to studying complex structural properties of social communities. In particular, the model allows us to study the influence of mechanisms related to individual and relational covariates on group formation processes.

The presented methodological developments aim to further our understanding of mechanisms driving the formation of social groups. First, they allow social scientists

Figure 7. Distribution of auxiliary statistics related to age in simulated partitions from models 1 and 2 for the first data set; (a) number of ties with given age differences; (b) intraclass correlation coefficient for age; (c) correlation between actors’ age and their group size.

Figure 8. Distribution of auxiliary statistics related to language in simulated partitions from models 1 and 2 for the second data set; (a) number of ties with a given language combination; (b) proportion of same-language ties; (c) average group size of a native German speaker.
to model and explain observations of self-assembled groups and potentially expand the range of social contexts that could be investigated. Moreover, some social processes widely studied at the dyadic level, such as homophily, can now be investigated at the group level. This modeling framework offers the possibility to explore different operationalizations of such mechanisms and assess which ones give a better representation of real-life processes. Finally, by moving from the dyad to the group perspective, mechanisms that have been suggested for group processes can be statistically tested.

Much remains to be discovered about this type of model. First, further research should address the issue of specifying such models. In particular, the theoretical implications of using particular covariate statistics (e.g., to represent homophilic processes) remain unknown. The comparability of statistics and parameters across different sizes of node sets is also a topic for future work. Second, a limitation of the presented framework is its inability to model observations of overlapping groups. As such groups are encountered in many social contexts, future research should extend the modeling framework to more general data representations, such as hypergraphs. Modeling group overlaps opens up the possibility of representing new dependencies between group memberships and analyzing, for example, what leads individuals to belong to multiple groups at the same time. Finally, the proposed model is cross-sectional in nature. Dynamic or longitudinal data offering rich insights on the processes driving social systems, an extension of this framework to a dynamic group representation would greatly further our understanding of social group dynamics.

APPENDIX A: PARTITION SETS WITH RESTRICTED GROUP SIZES: EXTENSION OF THE BELL NUMBERS AND STIRLING NUMBERS OF THE SECOND KIND

Extended Bell Numbers

The size of $\mathcal{P}$ is given by the Bell number $B_n$ (Bell 1934; Pitman 1997) and can be calculated iteratively by

$$B_0 = 1 \text{ and } B_{n+1} = \sum_{i=0}^{n} \binom{n}{i} B_i.$$  \hspace{1cm} (A1)

A similar formula can express the size of the space $\mathcal{P}'$ containing all partitions whose group sizes belong to the interval $[\sigma_{\min}, \sigma_{\max}]$.

To initialize the recurrence, we first know that the sets $\mathcal{P}'([1, n])$ are empty when $n$ is smaller then $\sigma_{\min}$. The minimal size is required for one correct partition to exist, therefore we have

$$B'_n = 0 \text{ for } 0 < n < \sigma_{\min} \text{ and } B'_{\sigma_{\min}} = 1.$$
partitions given by $B_i$. Here, we can enumerate the same partitions but the size $(n+1 - i)$ can only take values between $\sigma_{\text{min}}$ and $\sigma_{\text{max}}$, therefore $i$ can only vary from $i_{\text{min}}$ and $i_{\text{max}}$ defined as

$$i_{\text{min}} = \max(0, n+1 - \sigma_{\text{max}}) \quad \text{and} \quad i_{\text{max}} = \min(n, n+1 - \sigma_{\text{min}}). \quad (A2)$$

If we note $\mathcal{P}_i([[1, n+1]])$ the sets containing all partitions $P \in \mathcal{P}'([[1, n+1]])$ such that node $(n+1)$ belongs to a group of size $(n+1 - i)$:

$$\mathcal{P}_i([[1, n+1]]) = \{ P \in \mathcal{P}'([[1, n+1]]) \mid \#g_P(n+1) = n+1 - i \}, \quad (A3)$$

we can write the following:

$$\mathcal{P}'([[1, n+1]]) = \bigcup_{i=i_{\text{min}}}^{i_{\text{max}}} \mathcal{P}_i([[1, n+1]]).$$

For partitions in $\mathcal{P}_i([[1, n+1]])$, we first know there are $\binom{n}{i}$ ways to choose the group of $(n+1)$ and $B_i'$ ways to choose how to arrange the remaining $i$ nodes. This is true for any $i$ except when $(n+1 - i)$ corresponds to the whole set size $(n+1)$ (i.e., $i=0$). In that case, we use for convenience $B_0' = 1$.

We can therefore write $B_{n+1}'$ as the sum

$$B_{n+1}' = \sum_{i=i_{\text{min}}}^{i_{\text{max}}} \#\mathcal{P}_i([[1, n+1]]) = \sum_{i=i_{\text{min}}}^{i_{\text{max}}} \binom{n}{i} B_i'$$

and this establishes the recursive relation for $n \geq \sigma_{\text{min}}$.

**Extended Stirling Numbers**

The Stirling number $\binom{n}{m}$ is the number of partitions of $n$ nodes in $m$ blocks. Its calculation follows the relations (Nielsen 1906):

$$\binom{0}{0} = 1, \binom{n}{0} = \binom{n}{n} = 0 \quad \text{for } n>0, \binom{n+1}{m+1} = \sum_{i=m}^{n} \binom{n}{i} \binom{i}{m} \quad \text{for } m>0. \quad (A4)$$

Similarly, we can calculate $\psi_{\sigma_{\text{min}}, \sigma_{\text{max}}}(n,m)$, the number of partitions in $m$ blocks when blocks sizes belong to $[[\sigma_{\text{min}}, \sigma_{\text{max}}]]$. First, there is no possible partition for $n < m\sigma_{\text{min}}$, therefore

$$\psi_{\sigma_{\text{min}}, \sigma_{\text{max}}}(n,m) = 0 \quad \text{for } 0 < n < m\sigma_{\text{min}}.$$ 

The recursion then starts when there can be $m$ blocks of minimal size (i.e., $n = m\sigma_{\text{min}}$). To count possible partitions in this case, we first order all nodes in $n!$ different ways, and take each time the first group of $\sigma_{\text{min}}$ nodes, then the second group, and so on. Some partitions are counted several times, as we have $m$ possible ways to order these groups and $\sigma_{\text{min}}!$ ways to order the nodes inside each group. The final count is

$$\psi_{\sigma_{\text{min}}, \sigma_{\text{max}}}(n,m) = \frac{n!}{m!(\sigma_{\text{min}})!} \quad \text{for } n = m\sigma_{\text{min}}.$$
The terms in the sum of equation (A4) are the numbers of partitions where the node \((n+1)\) is in a group of size \((n+1)/C_i\) and the \(i\) remaining nodes are partitioned in \(m\) groups. As for \(B_n'\) numbers, we can adapt the original recursive relation with the indexes (equation A2):

\[
\psi_{a_{\text{min}}, a_{\text{max}}}(n+1, m+1) = \sum_{i = \text{min}}^{\text{max}} \binom{n}{i} \psi_{a_{\text{min}}, a_{\text{max}}}(i, m) \quad \text{for } n \geq m a_{\text{min}}.
\]

Finally, for the extreme case when the group of node \((n+1)\) is of size \((n+1)\) (i.e., \(i = 0\)) and there are no left groups to form (i.e., \(m = 0\)), we have to set for convenience:

\[
\psi_{a_{\text{min}}, a_{\text{max}}}(0, 0) = 1.
\]

**APPENDIX B: TRANSLATION OF THE EWENS DISTRIBUTION**

In this section, we demonstrate that the Ewens distribution (Ewens 1972), as defined by equation (3), can be expressed in the form of the exponential family introduced in this article with the following definition:

\[
Pr_{\lambda}(P = p) = \exp \left( \log (\lambda) \cdot \#p + \sum_{G \in P} \log (\#G-1)! \right) \frac{\kappa_{\theta}(\log (\lambda), 1)}{\kappa_{\theta}(\log (\lambda), 1)}. \tag{B1}
\]

To prove that this definition is equivalent to equation (3), we develop its numerator and denominator. Following properties of the exponential and logarithm functions, the numerator can be expressed for any partition \(P\):

\[
\exp (\log (\lambda) \cdot \#p + \sum_{G \in P} \log (\#G-1)! = \lambda^{\#p} \prod_{G \in P} \Gamma (\#G-1)! \tag{B2}
\]

Once this is established, proving the equivalence of the two definitions requires us to prove that the normalizing constant of our model simplifies to

\[
\kappa_{\theta}(\log (\lambda), 1) = \frac{\Gamma(\lambda+1)}{\Gamma(\lambda-1)} \tag{B3}
\]

Proving equation (B3) can be achieved by induction on \(n\) the number of nodes. The distribution (equation B1) is defined for statistics that are functions of the group sizes. We can therefore define the sequence \(\kappa_n\) relations found in equations (D4) and (D6) in Appendix D.

From equation (D4) and the property \(\Gamma(1) = 1\), we have for the basic case \(n = 1\):

\[
\kappa_1 = \exp (\log (\lambda) + \log (1)) = \lambda.
\]

We know from properties of the gamma function that \(\Gamma(\lambda+1) = \lambda \Gamma(\lambda)\), so we can validate the relation (equation B3) for \(n = 1\):

\[
\kappa_1 = \frac{\Gamma(\lambda)}{\Gamma(\lambda-1)}.
\]

Let us now use the previously demonstrated equation (D6) for higher values of \(n\):
\[ \kappa_{n+1} = \sum_{i=0}^{n} \binom{n}{i} \exp(\log(\lambda) + \log((n-i)!))\kappa_i = \sum_{i=0}^{n} \binom{n}{i} \lambda(n-i)! \kappa_i. \] (B4)

We then separate this sum into two parts, one containing the term corresponding to \(i=n\) and one containing the other terms:

\[ \kappa_{n+1} = \lambda \kappa_n + \sum_{i=0}^{n-1} \binom{n}{i} \lambda(n-i)! \kappa_i. \]

Finally, we develop the binomial coefficients and re-arrange them to find the definition of \(\kappa_n\) corresponding to the definition (equation B4) for \((n+1)\):

\[ \kappa_{n+1} = \lambda \kappa_n + \sum_{i=0}^{n-1} \frac{(n-1)!}{i!(n-1-i)!} \frac{n}{(n-i)!} \lambda(n-i)! \kappa_i \]
\[ = \lambda \kappa_n + n \left( \sum_{i=0}^{n-1} \binom{n-1}{i} \lambda((n-1)-i)! \kappa_i \right) \]
\[ = (\lambda + n) \kappa_n \]

If we assume that equation (B3) holds for a given \(n>0\), we therefore also have for \((n+1)\):

\[ \kappa_{n+1} = (\lambda + n) \frac{\Gamma(n+\lambda-1)}{\Gamma(\lambda-1)} = \frac{\Gamma((n+1)+\lambda-1)}{\Gamma(\lambda-1)}. \]

This proves that the relation (equation B3) holds for any integer \(n\) and that the model defined by equation (B1) is the same as the Ewens model defined by equation (3).

**APPENDIX C: INDEPENDENCE PROPERTIES OF THE DISTRIBUTION**

**Consistency**

Let \(P\) be a random partition over \(\mathcal{A}\) following equation (1) and \(P'\) a random partition over \(\mathcal{A}'\) following the same distribution with identical sufficient statistics and parameters. We pose \(\pi(P)\) the projection of \(P\) over the subset \(\mathcal{A}'\), and \(\pi^{-1}(P')\) the set of partitions over the node set \(\mathcal{A}\) whose projection is \(P'\).

Consistency of the distribution then implies equality between the marginal distribution of the random partition \(P\) over \(\mathcal{A}'\) and the distribution of \(P'\). In other words, the family of projections of a partition model on \(\mathcal{A}\) on the subset \(\mathcal{A}'\) is a partition model with the same sufficient statistics and same parameters. This property translates to:

\[ \Pr_\alpha(P \in \pi^{-1}(p')) = \Pr_\alpha(P' = p'). \]
Here we present some counter-examples of distributions used in this article for which this property does not hold. Let us use the space $\mathcal{A} = \{1, 2, 3\}$, its subset $\mathcal{A}' = \{1, 2\}$, and the projection $\pi$ from $\mathcal{P}(\mathcal{A})$ to $\mathcal{P}(\mathcal{A}')$.

**Uniform Model.** Let us use the uniform distribution over $\mathcal{P}(\mathcal{A})$. There are five different ways of partitioning this set, hence $1/5$ is the probability of any of these partitions. If we take a partition $\mathcal{P}' = \{\{1, 2\}\}$, we can calculate the marginal probability:

$$\Pr(\mathcal{P} \in \pi^{-1}(\mathcal{P}')) = \Pr(\mathcal{P} = \{\{1, 2, 3\}\}) + \Pr(\mathcal{P} = \{\{1, 2\}, \{3\}\}) = \frac{2}{5}$$

and the probability of observing $\{\{1, 2\}\}$ over $\mathcal{A}'$:

$$\Pr(\mathcal{P} = \mathcal{P}') = \Pr(\mathcal{P}' = \{\{1, 2\}\}) = \frac{1}{5}.$$

The uniform distribution is therefore not consistent.

**Model With One Statistic** $s_1(P) = \#P$. We can again use the same example on the same sets and $\mathcal{P}' = \{\{1, 2\}\}$. We have the following as marginal probability:

$$\Pr_{\alpha_1}(\mathcal{P} \in \pi^{-1}(\mathcal{P}')) = \Pr_{\alpha_1}(\mathcal{P} = \{\{1, 2, 3\}\}) + \Pr_{\alpha_1}(\mathcal{P} = \{\{1, 2\}, \{3\}\})$$

$$= \frac{\exp(\alpha_1) + \exp(2\alpha_1)}{\exp(\alpha_1) + 3 \exp(2\alpha_1) + \exp(3\alpha_1)}$$

and

$$\Pr_{\alpha_1}(\mathcal{P}' = \mathcal{P}') = \Pr_{\alpha_1}(\mathcal{P}' = \{\{1, 2\}\}) = \frac{\exp(\alpha_1)}{\exp(\alpha_1) + \exp(2\alpha_1)}.$$

Having these two terms equal is equivalent to the equation $\exp(2\alpha_1) = 0$, which has no solution in $\mathbb{R}$. Again, the consistency condition cannot be fulfilled for such models.

**Neutrality**

We now show that the neutrality property defined by equation (5) holds for any model defined for a set of statistics of the form

$$s_k(P) = \sum_{G \in \mathcal{P}} f_k(G),$$

with $(f_k)$ defined as real functions of the groups in the partition (i.e., representing any characteristic of the group). This definition covers all statistics used in this article; however, other types of statistics could also lead to neutral distributions.

Let $\mathcal{P}$ be a random partition defined for such a model with parameter vector $\alpha$. Furthermore, let $\mathcal{P}$ be the observed partition that has the property of being the union of its projections over the subsets $\mathcal{A}$ and $\mathcal{A}'$. We can write

$$\Pr_\alpha(\mathcal{P} = \mathcal{P} \mid \mathcal{P} = \pi(\mathcal{P}) \cup \pi'(\mathcal{P})) = \frac{\Pr_\alpha(\mathcal{P} = \mathcal{P} \mid \pi(\mathcal{P} \cup \pi'(\mathcal{P})) \neq \mathcal{P})}{\Pr_\alpha(\mathcal{P} = \pi(\mathcal{P} \cup \pi'(\mathcal{P})) \mid \mathcal{P} \neq \mathcal{P})}.$$
Because the observed partition verifies \( p = \pi(p) \cup \pi^c(p) \), the numerator simplifies to

\[
Pr_\alpha(P = p, P = \pi(p) \cup \pi^c(p)) = Pr_\alpha(P = p),
\]

and because summing over all groups of \( p \) is equivalent to summing over the groups in \( \pi(p) \) and \( \pi^c(p) \), this probability factorizes as follows:

\[
Pr_\alpha(P = \pi(P) \cup \pi^c(P)) = \frac{1}{\kappa_{\mathcal{P}(\mathcal{A})}(\alpha)} \exp \left( \sum_{k \in K} \alpha_k \left( \sum_{G \in \pi(p)} f_k(G) + \sum_{G \in \pi^c(p)} f_k(G) \right) \right) = \frac{1}{\kappa_{\mathcal{P}(\mathcal{A})}(\alpha)} \exp \left( \sum_{k \in K} \alpha_k \left( \sum_{G \in \pi(p)} f_k(G) + \sum_{G \in \pi^c(p)} f_k(G) \right) \exp \left( \sum_{k \in K} \alpha_k \sum_{G \in \pi^c(p)} f_k(G) \right) \right).
\]

(C1)

The denominator expresses the probability of having the random partition \( P = \pi(P) \cup \pi^c(P) \). It is the sum of probabilities of all partitions in \( \mathcal{P}(\mathcal{A}) \) with this property. If we define \( \mathcal{D}(\mathcal{A}, \mathcal{A}') \) as the set of these partitions, we can define a bijection \( b : \mathcal{D}(\mathcal{A}, \mathcal{A}') \rightarrow (\mathcal{P}(\mathcal{A}'), \mathcal{P}(\mathcal{A}'^c)) \) such that \( b(P) = (\pi_{\mathcal{A}'}(P), \pi_{\mathcal{A}'^c}(P)) \). We deduce:

\[
Pr_\alpha(P = \pi(P) \cup \pi^c(P)) = \frac{1}{\kappa_{\mathcal{P}(\mathcal{A})}(\alpha)} \exp \left( \sum_{k \in K} \alpha_k \left( \sum_{G \in \pi(p)} f_k(G) + \sum_{G \in \pi^c(p)} f_k(G) \right) \right) = \frac{1}{\kappa_{\mathcal{P}(\mathcal{A})}(\alpha)} \exp \left( \sum_{k \in K} \alpha_k \left( \sum_{G \in P_1} f_k(G) + \sum_{G \in P_2} f_k(G) \right) \right)\left( \sum_{k \in K} \alpha_k \sum_{G \in P_2} f_k(G) \right) = \frac{1}{\kappa_{\mathcal{P}(\mathcal{A})}(\alpha)} \left( \sum_{P_1 \in \mathcal{P}(\mathcal{A}')} \exp \left( \sum_{k \in K} \alpha_k \sum_{G \in P_1} f_k(G) \right) \right)\left( \sum_{P_2 \in \mathcal{P}(\mathcal{A}'^c)} \exp \left( \sum_{k \in K} \alpha_k \sum_{G \in P_2} f_k(G) \right) \right)
\]

and we can simplify:

\[
Pr_\alpha(P = \pi(P) \cup \pi^c(P)) = \frac{\kappa_{\mathcal{P}(\mathcal{A}')(\alpha)}\kappa_{\mathcal{P}(\mathcal{A}'^c)(\alpha)}}{\kappa_{\mathcal{P}(\mathcal{A})}(\alpha)}. \tag{C2}
\]

By dividing equations (C1) and (C2), the term \( \kappa_{\mathcal{P}(\mathcal{A})}(\alpha) \) simplifies and we finally have

\[
Pr_\alpha(P = p \mid P = \pi(P) \cup \pi^c(P)) = \frac{1}{\kappa_{\mathcal{P}(\mathcal{A})}(\alpha)} \exp \left( \sum_{k \in K} \alpha_k \sum_{G \in \pi(p)} f_k(G) \right) \times \frac{1}{\kappa_{\mathcal{P}(\mathcal{A}'^c)(\alpha)}(\alpha)} \exp \left( \sum_{k \in K} \alpha_k \sum_{G \in \pi^c(p)} f_k(G) \right) = Pr_\alpha(\pi(P) = \pi(p)) \times Pr_\alpha(\pi^c(P) = \pi^c(p))
\]

and this demonstrates the property of neutrality as defined by equation (5).
APPENDIX D: CALCULATION OF THE NORMALIZING CONSTANT WHEN STATISTICS ARE FUNCTIONS OF BLOCK SIZES

Here, we demonstrate that the normalizing constant $k$ as expressed by equation (2) can be calculated with a recursive formula when sufficient statistics ($s_k$) are of the form

$$s_k(P) = \sum_{G \in P} f_k(\#G),$$

with ($f_k$) defined as functions from the set of possible group sizes to $\mathbb{R}$. In the rest of the proof, we also pose

$$f(P) = \exp \left( \sum_{k \in K} \alpha_k \sum_{G \in P} f_k(\#G) \right).$$

In such cases, the probability distribution defined by equation (1) is said to be exchangeable (McCullagh 2011), as it is invariant under any permutation of the nodes. The normalizing constant $k$ then only depends on $n$ and is noted $k_n$.

For the sake of conciseness, we derive the recurrence relation that can be used to compute the constant $k_n'$ defined over the set $\mathcal{P}([[1,n]])$ with groups sizes between $\sigma_{\text{min}}$ and $\sigma_{\text{max}}$. The relation (equation 8) shown in the text for $k_n$ in the general case directly follows. Again, models defined for statistics of the form equation (7) on the set $\mathcal{P}$ are exchangeable, and we can write the constant $k_\mathcal{P}$ as $k_n'$ as it depends only on the number of nodes.

By using again the values $l_{\text{min}} = \max(0, n + 1 - \sigma_{\text{max}})$ and $l_{\text{max}} = \min(n, n + 1 - \sigma_{\text{min}})$, we can construct the sequence $k_n'$ with the following recursion:

$$k_n' = 0 \text{ for } 0 < n < \sigma_{\text{min}},$$

$$k_0' = 1, \quad k_{\sigma_{\text{min}}} = \exp \left( \sum_{k \in K} \alpha_k f_k(\sigma_{\text{min}}) \right),$$

$$k_{n+1}' = \sum_{i = l_{\text{min}}}^{l_{\text{max}}} \binom{n}{i} \exp \left( \sum_{k \in K} \alpha_k f_k(n + 1 - i) \right) k_i' \text{ for } n \geq \sigma_{\text{min}}.$$

The proof is based on a similar logic to the one used in Appendix A. To initialize the recursion, we know there are no possible partitions for smaller sizes, and there is only one partition with one group for $n = \sigma_{\text{min}}$. Therefore,

$$k_n' = 0 \text{ for } n < \sigma_{\text{min}},$$

$$k_n' = \exp \left( \sum_{k \in K} \alpha_k f_k(\sigma_{\text{min}}) \right) \text{ for } n = \sigma_{\text{min}}.$$

For $n > \sigma_{\text{min}}$, we can use the subsets $\mathcal{P}([[1,n+1]])$ defined by equation (A3) and write

$$k_{n+1} = \sum_{\bar{P} \in \mathcal{P}([[1,n+1]])} f(\bar{P}) = \sum_{i = l_{\text{min}}}^{l_{\text{max}}} \binom{n}{i} \sum_{\bar{P} \in \mathcal{P}([[1,n+1]])} f(\bar{P}).$$
Let us define \( G_i([1, n+1]) \) as the set of all possible groups of nodes in \([1, n+1]\) that include the node \((n+1)\) and whose size is equal to \((n+1 - i)\). To enumerate all possible partitions of \( P_i([1, n+1]) \), we enumerate all groups in \( G_i([1, n+1]) \) and all possible partitions over the remaining \( i \) nodes. With this notation, we have

\[
\kappa'_{n+1} = \sum_{i=i_{\text{min}}}^{i_{\text{max}}} \left( \sum_{g \in G_i([1, n+1])} \sum_{\tilde{P} \in P([1, n+1]) \setminus g} f(\tilde{P} \cup g) \right).
\]

Because the definition of the function \( f \) is invariant under permutations of the nodes, we can re-order the \( i \) remaining nodes from 1 to \( i \). From this, we deduce that for any group \( g \in G_i([1, n+1]) \) there is a bijection \( b_g : P([1, n+1]) \setminus g \rightarrow P([1, i]) \) such that partitions over remaining nodes are defined for these reordered nodes. We can therefore replace the sum indices in the previous expression:

\[
\kappa'_{n+1} = \sum_{i=i_{\text{min}}}^{i_{\text{max}}} \left( \sum_{g \in G_i([1, n+1])} \sum_{\tilde{P} \in P([1, i])} f(\tilde{P}) \right).
\]

We can use the definition (D1) of the statistics \( s_k \) to derive

\[
f(\tilde{P} \cup g) = \exp \left( \sum_{k \in K} \alpha_k \left( \sum_{G \in P} f_k(\#G) \exp \left( \sum_{k \in K} \alpha_k f_k(\#g) \right) \right) = f(\tilde{P}) f(g)
\]

and factorize

\[
\kappa'_{n+1} = \sum_{i=i_{\text{min}}}^{i_{\text{max}}} \left( \sum_{g \in G_i([1, n+1])} f(g) \sum_{\tilde{P} \in P([1, i])} f(\tilde{P}) \right).
\]

By definition, the following term simplifies to one of the previously calculated normalizing constants:

\[
\sum_{\tilde{P} \in P([1, i])} f(\tilde{P}) = \kappa'_i,
\]

except in the case of \( i=0 \) for which we set

\[
\kappa'_0 = 1. \quad (D5)
\]

Moreover, we know that for any \( g \in G_i([1, n+1]) \), \( f_k(\#g) = f_k(n+1 - i) \). Developing \( f(g) \) then removes any term depending on \( g \). The size of \( G_i([1, n+1]) \) being the number of ways to choose \( n - i \) elements (or \( i \) elements) among \( n \) nodes, we deduce

\[
\kappa'_{n+1} = \sum_{i=i_{\text{min}}}^{i_{\text{max}}} \binom{n}{i} \exp \left( \sum_{k \in K} \alpha_k f_k(n+1 - i) \right) \kappa'_i. \quad (D6)
\]
These expressions show that we can recursively construct the sequence $k'_n$, using the initialization (equation D4) and the recursive relation (equation D6). Given that the relation is linear and its factors are easy to calculate for a reasonable number of nodes, these normalizing constants can be directly calculated.

**APPENDIX E: PSEUODOCODE FOR THE ESTIMATION ALGORITHM**

Algorithm 1 shows the Metropolis-Hastings algorithm described in the section “Sampling Partitions” used to sample a set of $M$ partitions according to a given model. The algorithm begins from a starting partition $P_{\text{start}}$. The lengths of burn-in and thinning are represented, respectively, by the variables $b$ and $t$. The functions `get_neighbor` and `calculate_acceptance_ratio` are used for simplification. The first one picks a proposed partition, as a randomly picked neighbor partition from the initial partition. The neighbors of the initial partition are defined by the relation $R$, which can be any of the relations described in the section “Relations between Partitions,” or a combination of them. The second function calculates the acceptance ratio defined in the section “Sampling Partitions.”

Algorithm 2 describes the estimation procedure presented in the “Estimation Procedure” to obtain parameter estimates $\alpha_{\text{end}}$, as well as standard errors $se$ for inferential tests and a convergence ratio vector $c$. Again, some functions are used for simplification, including the `sample_partitions` function that was essentially described by algorithm 1. The other functions are `compute_statistics` (to calculate the sufficient statistics of the model for a given partition, as described in the section “Sufficient Statistics”), `calculate_covariance` (to calculate the covariance matrix between statistics in a partition sample), `calculate_convergence_ratios` (to compute the convergence vector $c$, as described in the section “Estimation Procedure”), and `calculate_standard_errors` (to compute the standard errors $se$, as described in the section “Estimation Procedure”).

Several details were intentionally simplified or removed from this pseudocode. Full description of the code can be found in the documentation provided in the supplementary material or at http://github.com/marion-hoffman/erpm.
Algorithm 1: Metropolis-Hastings algorithm to sample partitions

Data: Model statistics \( s \), model parameters \( \alpha \), number of sampled partitions \( M \), starting partition \( P_{\text{start}} \), relation \( R \), burn-in \( b \), thinning \( t \)

Result: List of sampled partitions \( \{P_i\}_{i=1}^M \)

\[ P_{\text{current}} \leftarrow P_{\text{start}}; \]
\[ \text{cpt} \leftarrow 1; \]
\[ \text{while } \text{cpt} \leq b + tM \text{ do} \]
\[ P_{\text{proposed}} \leftarrow \text{pick_neighbor}(R, P_{\text{current}}); \]
\[ \text{ratio} \leftarrow \text{calculate_acceptance_ratio}(R, P_{\text{current}}, P_{\text{proposed}}, s, \alpha); \]
\[ \text{rnd} \leftarrow \text{generate_random_number}(\text{min} = 0, \text{max} = 1); \]
\[ \text{if } \text{rnd} \leq \text{ratio} \text{ then} \]
\[ P_{\text{new}} \leftarrow P_{\text{proposed}}; \]
\[ \text{else} \]
\[ P_{\text{new}} \leftarrow P_{\text{current}}; \]
\[ \text{end} \]
\[ \text{if } \text{cpt} \geq b \text{ and } (\text{cpt} - b) \mod t = 0 \text{ then} \]
\[ i \leftarrow \frac{(\text{cpt} - b)}{t}; \]
\[ P_i \leftarrow P_{\text{new}}; \]
\[ \text{end} \]
\[ \text{cpt} \leftarrow \text{cpt} + 1; \]
\[ P_{\text{current}} \leftarrow P_{\text{new}}; \]
\[ \text{end} \]

Algorithm 2: Estimation procedure

Data: Model statistics \( s \), number of sampled partitions \( M_1 \) in phase 1 and \( M_2 \) in phase 2, scaling parameter \( \text{scaling} \), model statistics \( s \), starting parameters \( \alpha_{\text{start}}, \ldots \)

Result: Parameters \( \alpha_{\text{end}} \), standard errors \( \text{se} \), convergence ratios \( c \)

\[ s_{\text{obs}} \leftarrow \text{compute_statistics}(s, \text{Pobs}); \]

Phase 1:
\[ \text{samplephase1} \leftarrow \text{sample_partitions}(s, \alpha_{\text{start}}, M_1, \ldots); \]
\[ \text{cov_matrix} \leftarrow \text{calculate_covariance}(	ext{samplephase1}); \]
\[ D_0 \leftarrow \text{cov_matrix}; \]
\[ \text{non_diagonal_elements}(D_0) \leftarrow \text{scaling} \ast \text{non_diagonal_elements}(D_0); \]
\[ s_{\text{phase1}} \leftarrow \text{mean}(	ext{compute_statistics}(s, \text{samplephase1})); \]
\[ \alpha_{\text{start}} \leftarrow \alpha_{\text{start}} - a \ast D_0^{-1} \ast (s_{\text{phase1}} - s_{\text{obs}}); \]

Phase 2:
\[ \alpha_{1,1} \leftarrow \alpha_{\text{start}}; \]
\[ \text{for } r \leftarrow 1 \text{ to } R \text{ do} \]
\[ a_r \leftarrow \frac{a}{2^{(ar/3)}}; \]
\[ N \leftarrow 1; \]
\[ \text{while } N \geq 2^{(ar/3)} \text{ and } N \leq 2^{(ar/3) + 200} \text{ do} \]
\[ \text{samplephase2}, r, N \leftarrow \text{sample_partitions}(s, \alpha_{r,N}, 1, \ldots); \]
\[ s_{\text{phase2}}, r, N \leftarrow \text{compute_statistics}(s, \text{samplephase2}, r, N); \]
\[ \alpha_{r,N+1} \leftarrow \alpha_{r,N} - a_r \ast D_0^{-1} \ast (s_{\text{phase2}}, r, N - s_{\text{obs}}); \]
\[ \text{if } N \geq 2^{(ar/3)} \text{ and } \text{sampled} \text{ statistics crossed } s_{\text{obs}} \text{ then} \]
\[ \text{break}; \]
\[ \text{if } r < R \text{ then} \]
\[ \alpha_{r+1,1} \leftarrow \text{mean}(\alpha_r); \]
\[ \text{else} \]
\[ \alpha_{\text{end}} \leftarrow \text{mean}(\alpha_R); \]

Phase 3
\[ \text{samplephase3} \leftarrow \text{sample_partitions}(s, \alpha_{\text{end}}, M_3, \ldots); \]
\[ c \leftarrow \text{calculate_convergence_ratios}(	ext{samplephase3}); \]
\[ \text{cov_matrix} \leftarrow \text{calculate_covariance}(	ext{samplephase3}); \]
\[ \text{se} \leftarrow \text{calculate_standard_errors}(	ext{cov_matrix}); \]
Acknowledgments

We thank the members of the Chair of Social Networks at ETH Zürich and the members of the Duisterbelt for useful comments and feedback.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Marion Hoffman acknowledges IAST funding from the French National Research Agency under the Investments for the Future (Investissements d’Avenir) program (grant ANR-17-EURE-0010). Per Block is funded by an Eccellenza grant of the Swiss National Science Foundation and supported by the Leverhulme Centre for Demographic Science.

ORCID iDs

Marion Hoffman https://orcid.org/0000-0002-0741-7760
Per Block https://orcid.org/0000-0002-7583-2392

Supplemental Material

Supplemental material for this article is available online.

References

Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew. 1954. The Nature of Prejudice. Cambridge, MA: Addison-Wesley.
Antoniak, Charles E. 1974. “Mixtures of Dirichlet Processes with Applications to Bayesian Non-parametric Problems.” Annals of Statistics 2:1152–74.
Bailey, Jasmonitor L., and John Skvoretz. 2017. “The Social-Psychological Aspects of Team Formation: New Avenues for Research.” Sociology Compass 11:1–12.
Balest, Jessica, Laura Secco, Elena Pisani, and Alberto Caimo. 2019. “Sustainable Energy Governance in South Tyrol (Italy): A Probabilistic Bipartite Network Model.” Journal of Cleaner Production 221: 854–62.
Bell, Eric Temple. 1934. “Exponential Polynomials.” Annals of Mathematics 2:258–77.
Besag, Julian. 1974. “Spatial Interaction and the Statistical Analysis of Lattice Systems.” Journal of the Royal Statistical Society: Series B (Statistical Methodology) 36:192–225.
Blau, Peter Michael. 1977. Inequality and Heterogeneity: A Primitive Theory of Social Structure. Vol. 7. New York: Free Press.
Block, P. 2018. “Network Evolution and Social Situations.” Sociological Science 5:402–31.
Bomiriya, Rashmi Pankajai. 2014. “Topics in Exponential Random Graph Modeling.” PhD dissertation, Pennsylvania State University, State College.
Briscoe, Gerard, and Catherine Mulligan. 2014. “Digital Innovation: The Hackathon Phenomenon.” Working Paper No. 6. London: Creative Works.
Butts, Carter T. 2007. “Models for Generalized Location Systems.” Sociological Methodology 37: 283–348.
Butts, Carter T. 2022. “A Dynamic Process Reference Model for Sparse Networks with Reciprocity.” Journal of Mathematical Sociology 46:1–27.
Butts, Carter T., and Zack W. Almquist. 2015. “A Flexible Parameterization for Baseline Mean Degree in Multiple-Network ERGMs.” Journal of Mathematical Sociology 39:163–67.
Cappé, Olivier, Eric Moulines, and Tobias Rydén. 2005. Inference in Hidden Markov Models. Berlin, Germany: Springer-Verlag.
Casciari, Tiziana, and Miguel Sousa Lobo. 2008. “When Competence Is Irrelevant: The Role of Interpersonal Affect in Task-Related Ties.” Administrative Science Quarterly 53:655–84.
Chib, Siddhartha, and Edward Greenberg. 1995. “Understanding the Metropolis-Hastings Algorithm.” *American Statistician* 49:327–35.

Connor, Robert J., and James E. Mosimann. 1969. “Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution.” *Journal of the American Statistical Association* 64: 194–206.

Contractor, Noshir. 2013. “Some Assembly Required: Leveraging Web Science to Understand and Enable Team Assembly.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371:1–14.

Crane, Harry. 2016. “The Ubiquitous Ewens Sampling Formula.” *Statistical Science* 31:1–19.

Creel, Scott, and Nancy Marusha Creel. 1995. “Communal Hunting and Pack Size in African Wild Dogs, Lycaon Pictus.” *Animal Behaviour* 50:1325–39.

Deuflhard, Peter. 2011. *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms.* New York: Springer.

Ewens, Warren J. 1972. “The Sampling Theory of Selectively Neutral Alleles.” *Theoretical Population Biology* 3:87–112.

Falk-Krzesinski, Holly J., Katy Börner, Noshir Contractor, Stephen M. Fiore, Kara L. Hall, Joann Keyton, Bonnie Spring, Daniel Stokols, William Trochim, and Brian Uzzi. 2010. “Advancing the Science of Team Science.” *Clinical and Translational Science* 3:263–66.

Feld, Scott L. 1982. “Social Structural Determinants of Similarity among Associates.” *American Sociological Review* 47:797–801.

Ferguson, Thomas S. 1973. “A Bayesian Analysis of Some Nonparametric Problems.” *Annals of Statistics* 1:209–230.

Fischer, Claude S. 1982. *To Dwell among Friends: Personal Networks in Town and City.* Chicago: University of Chicago Press.

Frank, Ove, and David Strauss. 1986. “Markov Graphs.” *Journal of the American Statistical Association* 81:832–42.

Geiger, Dan, and David Heckerman. 1997. “A Characterization of the Dirichlet Distribution through Global and Local Parameter Independence.” *Annals of Statistics* 25:1344–69.

Gelman, Andrew, and Xiao-Li Meng. 1998. “Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling.” *Statistical Science* 13:163–85.

Geyer, Charles J., and Elizabeth A. Thompson. 1995. “Constrained Monte Carlo Maximum Likelihood for Dependent Data.” *Journal of the Royal Statistical Society: Series B (Methodological)* 54:657–83.

Gittleman, John L. 1989. “Carnivore Group Living: Comparative Trends.” Pp. 183–207 in *Carnivore Behavior, Ecology, and Evolution*, edited by J. L Gittleman. Boston: Springer.

Gómez-Zarà, Diego, Matthew Paras, Marlon Twyman, Jacqueline N. Lane, Leslie Ann DeChurch, and Noshir Contractor. 2019. “Who Would You Like to Work With?” Pp. 1–15 in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York: Association for Computing Machinery.

Gompers, Paul A., Kevin Huang, and Sophie Q. Wang. 2017. “Homophily in Entrepreneurial Team Formation.” NBER Working Paper No. 23459. Cambridge, MA: National Bureau of Economic Research.

Granovetter, Mark. 1985. “Economic Action and Social Structure: The Problem of Embeddedness.” *American Journal of Sociology* 91:481–510.

Gu, Ming Gao, and Fan Hui Kong. 1998. “A Stochastic Approximation Algorithm with Markov Chain Monte-Carlo Method for Incomplete Data Estimation Problems.” *Proceedings of the National Academy of Sciences* 95:7270–74.

Gu, Ming Gao, and Hong-Tu Zhu. 2001. “Maximum Likelihood Estimation for Spatial Models by Markov Chain Monte Carlo Stochastic Approximation.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63:339–55.

Guimerà, Roger, Brian Uzzi, Jarrett Spiro, and Luis A. Nunes Amaral. 2005. “Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance.” *Science* 308:697–702.
Hoffman et al. 39

Hammersley, John M., and Peter Clifford. 1971. “Markov Fields on Finite Graphs and Lattices.” Unpublished manuscript.

Handcock, Mark S. 2003. “Assessing Degeneracy in Statistical Models of Social Networks.” Working Paper No. 39. Seattle: Center for Statistics and the Social Sciences, University of Washington.

Hastings, W. Keith. 1970. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications.” *Biometrika* 57:97–109.

Hintze, Jerry L., and Ray D. Nelson. 1998. “Violin Plots: A Box Plot-Density Trace Synergism.” *American Statistician* 52:181–84.

Hoffman, Marion, Per Block, Timon Elmer, and Christoph Stadtfeld. 2020. “A Model for the Dynamics of Face-to-Face Interactions in Social Groups.” *Network Science* 8:1–22.

Homans, George C. 1950. *The Human Group*. New York: Harcourt.

Hubbell, Stephen P. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ: Princeton University Press.

Hummel, Ruth M., David R. Hunter, and Mark S. Handcock. 2012. “Improving Simulation-Based Algorithms for Fitting ERGMs.” *Journal of Computational and Graphical Statistics* 21:920–39.

Hunter, David R. 2007. “Curved Exponential Family Models for Social Networks.” *Social Networks* 29:216–30.

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481), 248–258.

Hunter, David R., and Mark S. Handcock. 2006. “Inference in Curved Exponential Family Models for Networks.” *Journal of Computational and Graphical Statistics* 15:565–83.

Kalleberg, Arne L., David Knopke, Peter V. Marsden, and Joe L. Spaeth. 1996. *Organizations in America: Analysing Their Structures and Human Resource Practices*. Thousand Oaks, CA: Sage.

Kingman, John Frank Charles. 1978. “Random Partitions in Population Genetics.” *Proceedings of the Royal Society of London A: Mathematical and Physical Sciences* 361:1–20.

Krackhardt, David. 1988. “Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data.” *Social Networks* 10:359–81.

Krivitsky, Pavel N., Mark S. Handcock, and Martina Morris. 2011. “Adjusting for Network Size and Composition Effects in Exponential-Family Random Graph Models.” *Statistical Methodology* 8:319–39.

Lai, Tze Leung. 2003. “Stochastic Approximation.” *Annals of Statistics* 31:391–406.

Lara, Miguel, and Kate Lockwood. 2016. “Hackathons as Community-Based Learning: A Case Study.” *TechTrends* 60:486–95.

Lauritzen, Steffen L. 1996. *Graphical Models*. Oxford, UK: Clarendon.

Lazarsfeld, Paul F., and Robert K. Merton. 1954. “Friendship as a Social Process: A Substantive and Methodological Analysis.” Pp. 8–66 in *Freedom and Control in Modern Society*, edited by M. Berger. New York: Van Nostrand.

Lewin, Kurt, Fritz Trans Heider, and Grace M. Heider. 1936. *Principles of Topological Psychology*. New York: McGraw-Hill.

Lindenberg, Siegwart. 1997. “Grounding Groups in Theory: Functional, Cognitive, and Structural Interdependencies.” *Advances in Group Processes* 14:281–331.

Lungeanu, Alina, Dorothy R. Carter, Leslie A. DeChurch, and Noshir S. Contractor. 2018. “How Team Interlock Ecosystems Shape the Assembly of Scientific Teams: A Hypergraph Approach.” *Communication Methods and Measures* 12:174–98.

Lusher, Dean, Johan Koskinen, and Garry Robins. 2013. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge, UK: Cambridge University Press.

McCullagh, Peter. 2011. “Random Permutations and Partition Models.” Pp. 1170–77 in *International Encyclopedia of Statistical Science*, edited by M. Lovric. Berlin, Germany: Springer-Verlag.

McPherson, Miller, and Lynn Smith-Lovin. 1987. “Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups.” *American Sociological Review* 52:370–79.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27:415–44.
Snijders, Tom A. B. 2001. “The Statistical Evaluation of Social Network Dynamics.” *Sociological Methodology* 31:361–95.

Snijders, Tom A. B. 2002. “Markov Chain Monte Carlo Estimation of Exponential Random Graph Models.” *Journal of Social Structure* 3:1–40.

Snijders, Tom A. B., and Alessandro Lomi. 2019. “Beyond Homophily: Incorporating Actor Variables in Statistical Network Models.” *Network Science* 7:1–19.

Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. “New Specifications for Exponential Random Graph Models.” *Sociological Methodology* 36:99–153.

Sundberg, Rolf. 2019. *Statistical Modelling by Exponential Families*. Cambridge, UK: Cambridge University Press.

Topi, Heikki, and Allen Tucker. 2014. *Computing Handbook: Information Systems and Information Technology*. Boca Raton, FL: CRC Press.

Wang, Peng, Ken Sharpe, Garry L. Robins, and Philippa E. Pattison. 2009. “Exponential Random Graph (p*) Models for Affiliation Networks.” *Social Networks* 31:12–25.

Wasserman, Stanley, and Philippa Pattison. 1996. “Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and p*.” *Psychometrika* 61:401–25.

Wedderburn, Robert W. M. 1976. “On the Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models.” *Biometrika* 63:27–32.

Zhu, Mengxiao, Yun Huang, and Noshir S. Contractor. 2013. “Motivations for Self-Assembling into Project Teams.” *Social Networks* 35:251–64.

**Author Biographies**

**Marion Hoffman** is a research fellow at the Institute for Advanced Study in Toulouse. She was previously a PhD student at ETH Zürich. Her research aims at furthering statistical methods for the analysis of social networks and social groups. Her work has focused on the development of relational event models and exponential family models tailored for social group data.

**Per Block** is an SNSF Eccellenza Professor at the University of Zürich. Previously, he worked at the Leverhulme Centre for Demographic Science and the department of Sociology at the University of Oxford and at the ETH Zürich. His research area is the sociology of networks, with a focus on structural patterns that represent the interdependence of different observations, as well as the development of new methodology and furthering our understanding of the current statistical network toolbox.

**Tom A. B. Snijders** is a professor of statistics and methodology in the social sciences at the University of Groningen and an emeritus fellow of Nuffield College, University of Oxford. His research concentrates on multilevel analysis and social network analysis. His work on developing statistical methodology for network dynamics is implemented in the software SIENA (Simulation Inference for Empirical Network Analysis), which is available as the package RSiena in R.