

Article

Stochastic Simulations of Casual Groups

José F. Fontanari 

Instituto de Física de São Carlos, Universidade de São Paulo, P.O. Box 369, São Carlos 13560-970, SP, Brazil;
fontanari@ifsc.usp.br

Abstract: Free-forming or casual groups are groups in which individuals are in face-to-face interactions and are free to maintain or terminate contact with one another, such as clusters of people at a cocktail party, play groups in a children's playground or shopping groups in a mall. Stochastic models of casual groups assume that group sizes are the products of natural processes by which groups acquire and lose members. The size distributions predicted by these models have been the object of controversy since their derivation in the 1960s because of the neglect of fluctuations around the mean values of random variables that characterize a collection of groups. Here, we check the validity of these mean-field approximations using an exact stochastic simulation algorithm to study the processes of the acquisition and loss of group members. In addition, we consider the situation where the appeal of a group of size i to isolates is proportional to i^α . We find that, for $\alpha \leq 1$, the mean-field approximation fits the equilibrium simulation results very well, even for a relatively small population size N . However, for $\alpha > 1$, this approximation scheme fails to provide a coherent description of the distribution of group sizes. We find a discontinuous phase transition at $\alpha_c > 1$ that separates the regime where the variance of the group size does not depend on N from the regime where it grows linearly with N . In the latter regime, the system is composed of a single large group that coexists with a large number of isolates. Hence, the same underlying acquisition-and-loss process can explain the existence of small, temporary casual groups and of large, stable social groups.

Keywords: casual groups; size distribution; stochastic simulation; mean-field approximation; phase transition

MSC: 91Dxx; 92-10; 60Gxx



Citation: Fontanari, J.F. Stochastic Simulations of Casual Groups.

Mathematics **2023**, *11*, 2152. <https://doi.org/10.3390/math11092152>

Academic Editor: Takashi Suzuki

Received: 3 April 2023

Revised: 24 April 2023

Accepted: 2 May 2023

Published: 4 May 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Finding and explaining patterns in the ebb and flow of people in a public gathering have been challenging tasks for the mathematically inclined social scientist [1,2]. People constantly join and leave groups, so at any moment, the gathering appears as a collection of social clusters. The stochastic models proposed in the 1960s to explain the size distribution of these free-forming or casual groups ignore any prior knowledge of the individuals present in the gathering and use only a couple of parameters to represent people's average tendencies to join and leave groups [3,4]. Here, we evaluate the suitability of these models to reproduce the empirical size distribution of casual groups.

In fact, the equilibrium group-size distributions resulting from these null models, viz., the zero-truncated Poisson distribution and the logarithmic distribution, well describe the observed size distributions of collections of small groups, such as pedestrians on a sidewalk, playgroups in a playground and shopping groups [3]. However, these distributions were derived using a mean-field approximation to solve equations for the expected number of groups of a given size, which prompted the criticism that they were not actual outcomes of the models but artifacts of the approximation scheme [4–6]. Here, we attempt to settle this long-standing (and likely forgotten) issue using Gillespie's stochastic algorithm to exactly simulate the group dynamics [7,8].

In addition, we extend the null models proposed in the 1960s to consider the situation where the appeal of a group of size i to isolates (i.e., individuals who are not members of any group) is proportional to i^α , where $\alpha \in (-\infty, \infty)$ is the attractiveness exponent. Large negative values of α describe situations where a predominance of couples and isolates is expected, whereas large positive values of α foster the formation of a single large group that coexists with isolates. In time, the truncated Poisson distribution is obtained for $\alpha = 0$ [3], and the logarithmic distribution is obtained for $\alpha = 1$ [4], which have been the only cases studied so far. The stochastic simulations of the group dynamics indicate that the mean-field approximation yields the exact results for $\alpha \leq 1$ and a large population size N . However, the approximation fails for $\alpha > 1$ since it violates the fixed-population constraint.

The variation in the attractiveness exponent α allows the modeling of collections of small, temporary groups as well as of large, stable groups. We find that these two scenarios are separated by a discontinuous phase transition at $\alpha = \alpha_c > 1$. The probability of observing large groups vanishes exponentially with increasing group size i for $\alpha < \alpha_c$. For $\alpha > \alpha_c$, the probability mass function concentrates around $i = 1$ and $i = N - n_1$, where n_1 is the number of isolates. Hence, the acquisition-and-loss process of group dynamics does not produce a power-law decay for the probability of finding large groups observed in face-to-face interaction networks [9]. Long-tailed group-size distributions are outputs of agent-based models where the individuals are ascribed distinct degrees of attractiveness [10,11], so it seems that some knowledge of the individuals present in the gathering is necessary to produce these power-law distributions. In fact, the natural tendency of people to gravitate toward others who share similar interests or backgrounds is an important factor in explaining the formation of social groups [12–14]. Nevertheless, it is noteworthy that the stochastic null models can explain the empirical size distribution of collections of small groups.

A more recent and fruitful approach to the characterization of social groups or, more precisely, social networks—networks of friends or other acquaintances—is based on the complex networks framework [15]. In particular, a group of individuals (nodes) with a high density of internal links but with a comparatively lower density of external links is called a community [16,17]. Communities are ubiquitous in social and biological systems and are believed to represent real social groupings assembled by interest or background. We note that, although there are a variety of public repositories of animal social networks (see, e.g., [18,19]), the detection of communities in the real world as well as in artificial networks is a challenging computational task [20,21]. From an evolutionary perspective, the community organization of social networks is considered optimal if it boosts communication and decision making at the group level while keeping a minimum number of connections between individuals [22,23]. However, social network communities are relatively stable groups and thus are not good models for fleeting casual groups, which can be described by the less popular face-to-face interaction networks [24].

The rest of this paper is organized as follows: In Section 2, we describe the group dynamics and derive the exact equations for the expected number of groups of a given size. In Section 3, we present a brief overview of Gillespie’s algorithm and study the group dynamics using this stochastic simulation algorithm. In Section 4, we solve the equations derived in Section 2 for the equilibrium regime using the mean-field approximation and present explicit analytical expressions for the cases $\alpha \rightarrow -\infty$, $\alpha = 0$ and $\alpha = 1$, which are then compared with the simulation results. In Section 5, we study the equilibrium regime for $\alpha > 1$ using stochastic simulations and show that in the limit $N \rightarrow \infty$, there is a discontinuous phase transition separating the scenarios where the variance of the group size is finite and where it diverges linearly with increasing N . Finally, in Section 6, we review our main results and present some concluding remarks.

2. The Model

We consider a fixed number of individuals N that organize themselves into a variable number of groups of size $i = 1, \dots, N$ in a closed system. We denote by n_i the number of groups of size i at time t . These random variables satisfy the constraint $\sum_{i=1}^N i n_i = N$ and determine the total number of groups in the system, viz., $\sum_{i=1}^N n_i = M$, which is also a random variable. The processes of joining and leaving the groups are as follows.

Each individual in a group of size $i > 1$ has a probability $\mu\delta t$ of leaving the group during the time interval δt . When an individual leaves a group, it becomes an isolate, i.e., a group of size $i = 1$. An isolate has a probability $\lambda\delta t$ of joining a group (including other isolates) in the time interval δt .

We assume that the attractiveness of a group of size i to isolates is proportional to i^α . The case $\alpha = 0$ describes the situation where isolates join any group in the system at the same rate, whatever its size [3]. For $\alpha > 0$, we have a contagious scenario that favors the formation of large groups [4], whereas, for $\alpha < 0$, we have an aversion scenario that disfavors the formation of groups. In the 1960s, approximate analytical expressions for the expected values of n_i were derived for the cases $\alpha = 0$ and $\alpha = 1$ only [3,4].

The understanding of the processes by which groups acquire and lose individuals is facilitated if we write down the conditional expected values of the random variables $n_i(t + \delta t)$ given that the system is in the state $\vec{n}(t) = (n_1(t), \dots, n_N(t))$ at time t . Let us begin with the conditional expectation of the number of isolates,

$$\mathbb{E}[n_1(t + \delta t) | \vec{n}(t)] = n_1 + \mu\delta t(2n_2) + \mu\delta t(N - n_1) - \lambda\delta t n_1 \frac{N_\alpha - n_1}{N_\alpha - 1} - 2\lambda\delta t n_1 \frac{n_1 - 1}{N_\alpha - 1}, \quad (1)$$

where we have omitted the dependence of the variables $n_i(t)$ on t that appear on the right-hand side (RHS) of the equation. In addition, we have introduced the notation

$$N_\alpha(t) = 1^\alpha n_1(t) + 2^\alpha n_2(t) + \dots + N^\alpha n_N(t), \quad (2)$$

so that $N_0(t) = M(t)$ and $N_1(t) = N$. The third term on the RHS of Equation (1) takes into account the fact that any individual who is not isolated may become an isolate with a probability of $\mu\delta t$. The second term corrects the third term by reckoning with the fact that whenever an individual leaves a group of two individuals, two isolates are created. The fourth term on the RHS of Equation (1) accounts for the event that an isolate joins any group with $i > 1$ individuals, whereas the fifth term accounts for the aggregation of two isolates.

Next, we consider the conditional expectation of the number of couples, viz.,

$$\mathbb{E}[n_2(t + \delta t) | \vec{n}(t)] = n_2 + 3\mu\delta t n_3 - 2\mu\delta t n_2 - \lambda\delta t n_1 \frac{2^\alpha n_2}{N_\alpha - 1} + \lambda\delta t n_1 \frac{n_1 - 1}{N_\alpha - 1}. \quad (3)$$

The second and third terms on the RHS of Equation (3) account for the facts that a group of two individuals is created when an individual leaves a group of three individuals, and it is destroyed when an individual leaves a group of two individuals. The fourth term accounts for the event that isolates attracted by groups of two individuals produce groups of three individuals, and the fifth term accounts for the joining of two isolates to produce a group of two individuals.

For groups with $i = 3, \dots, N$ individuals, we can write the general expression

$$\mathbb{E}[n_i(t + \delta t) | \vec{n}(t)] = n_i + (i + 1)\mu\delta t n_{i+1} - i\mu\delta t n_i - \lambda\delta t n_1 \frac{i^\alpha n_i}{N_\alpha - 1} + \lambda\delta t n_1 \frac{(i - 1)^\alpha n_{i-1}}{N_\alpha - 1}, \quad (4)$$

with $n_{N+1} \equiv 0$. The interpretation of the terms in this equation follows straightforwardly from the interpretations of the terms in Equations (1) and (3).

We stress that Equations (1), (3) and (4) for the conditional expectations of $n_i(t + \delta t)$, $i = 1, \dots, N$ are exact. Adding these equations yields the conditional expectation for the number of groups,

$$\mathbb{E}[M(t + \delta t) | \vec{n}(t)] = M(t) + \mu \delta t [N - n_1(t)] - \lambda \delta t n_1(t), \quad (5)$$

from which we can see that the isolates play a key role in driving the casual group dynamics. Averaging Equations (1), (3)–(5) over the states $\vec{n}(t)$ and taking the limit $\delta t \rightarrow 0$ yield

$$\frac{d\langle n_1 \rangle}{dt} = 2\mu \langle n_2 \rangle + \mu(N - \langle n_1 \rangle) - \lambda \mathbb{E} \left[\frac{n_1(N_\alpha - n_1)}{N_\alpha - 1} \right] - 2\lambda \mathbb{E} \left[\frac{n_1(n_1 - 1)}{N_\alpha - 1} \right], \quad (6)$$

$$\frac{d\langle n_2 \rangle}{dt} = 3\mu \langle n_3 \rangle - 2\mu \langle n_2 \rangle - 2^\alpha \lambda \mathbb{E} \left[\frac{n_1 n_2}{N_\alpha - 1} \right] + \lambda \mathbb{E} \left[\frac{n_1(n_1 - 1)}{N_\alpha - 1} \right], \quad (7)$$

$$\frac{d\langle n_i \rangle}{dt} = (i + 1)\mu \langle n_{i+1} \rangle - i\mu \langle n_i \rangle - i^\alpha \lambda \mathbb{E} \left[\frac{n_1 n_i}{N_\alpha - 1} \right] + (i - 1)^\alpha \lambda \mathbb{E} \left[\frac{n_1 n_{i-1}}{N_\alpha - 1} \right] \quad (8)$$

for $i = 2, \dots, N$, and

$$\frac{d\langle M \rangle}{dt} = \mu N + (\lambda - \mu) \langle n_1 \rangle, \quad (9)$$

where $\langle n_{N+1} \rangle \equiv 0$, and we have introduced the notation $\langle n_i(t) \rangle \equiv \mathbb{E}[n_i(t)]$. We note that by rescaling the time $\tau = \mu t$, these equations depend on the aggregation and disaggregation rates only through their ratio,

$$\kappa \equiv \frac{\lambda}{\mu}. \quad (10)$$

Of course, Equations (6)–(8) do not form a closed set of equations since there are quantities (e.g., $\mathbb{E}[n_1^2 / (N_\alpha - 1)]$) that are left undefined. Somewhat surprisingly, however, in the equilibrium regime, i.e., $d\langle n_i \rangle / dt = 0$ for $i = 1, \dots, N$, Equation (9) yields the exact mean number of isolates,

$$\langle n_1 \rangle_{eq} = \frac{N}{1 + \kappa}, \quad (11)$$

which does not depend on the attractiveness exponent α .

3. The Gillespie Algorithm

Here, we offer a brief overview of Gillespie's algorithm for simulating continuous-time stochastic models [7,8]. In the time interval δt , the probability that aggregation occurs is $\lambda n_1(t) \delta t$, and the probability that an individual leaves a group is $\mu [N - n_1(t)] \delta t$. Since these two events decrease and increase the number of groups by one unity, respectively, their probabilities appear on the RHS of Equation (5). Given the state $\vec{n}(t)$ at time t , the probability that the next event will occur in the infinitesimal time interval $(t + t', t + t' + \delta t')$ is $P(t') \delta t'$, where $P(t')$ is the exponential distribution,

$$P(t') = v \exp(-t'v). \quad (12)$$

and

$$v = \mu N + (\lambda - \mu) n_1(t) \quad (13)$$

is the total rate of events. The event that occurs in the time interval $(t + t', t + t' + \delta t')$ is an aggregation with a probability $\lambda n_1(t) / v$ and a disaggregation with a probability $\mu [N - n_1(t)] / v$. In the case that aggregation occurs, there are two possibilities: an isolate can join a group of size $i > 1$, which is an event that happens with a probability

$$\frac{i^\alpha n_i(t)}{n_1(t) - 1 + 2^\alpha n_2(t) + \dots + N^\alpha n_N(t)}, \quad (14)$$

or two isolates can join together to form a couple, which has a probability

$$\frac{n_1(t) - 1}{n_1(t) - 1 + 2^\alpha n_2(t) + \dots + N^\alpha n_N(t)}. \quad (15)$$

In the case that disaggregation occurs in the time interval $(t + t', t + t' + \delta t')$, a single individual leaves a group of size $i > 1$, which is an event that happens with a probability

$$\frac{in_i(t)}{2n_2(t) + \dots + Nn_N(t)} = \frac{in_i(t)}{N - n_1(t)}. \quad (16)$$

In sum, given the state of the system $\vec{n}(t)$ at time t , the stochastic simulation of the casual-group model begins with the choice of the time t' when the next event will occur using the distribution in (12), followed by the choice of the type of event—aggregation with a probability $\lambda n_1(t)/v$ and disaggregation with a probability $\mu[N - n_1(t)]/v$. Finally, the specific aggregation and disaggregation events are chosen with probabilities given by Equations (14)–(16). We note that, with the exception of the determination of the time of the next event, the aggregation and disaggregation rates always appear in the form of the ratio λ/μ . This numerical algorithm produces the exact trajectories of the states $\vec{n}(t)$, which, when properly averaged over many independent runs, offers the only way to verify the validity of the approximation schemes used to solve high-dimensional master equations [7,8].

Here, we use the bracket notation $\langle \dots \rangle$ to indicate the average over independent runs. In fact, because the number of runs is very large (typically 10^6), we can safely equate the average over independent runs to the expected values of the random variables $n_1(t), \dots, n_N(t)$, hence the choice of the same bracket notation used in Section 2.

Figure 1 shows the time evolution of the mean density of isolates, $\langle n_1(t) \rangle / N$, and the mean number of individuals per group $m_\alpha \equiv \langle N/M(t) \rangle$ for different values of the attractiveness exponent α . At time $t = 0$, all N individuals are isolates (i.e., $n_1(0) = N$ and $n_i(0) = 0$, $i > 1$). Interestingly, although the mean density of isolates does not depend on α in the equilibrium regime, as shown in Equation (11), the transient regime is affected by the attractiveness exponent: the increase in α favors the production of isolates. This is so because $\alpha > 0$ hinders the formation of couples, which requires the annihilation of two isolates, whereas the formation of groups of size $i > 2$ requires the annihilation of only one isolate. In addition, an increase in α increases the transient period, as well as the mean size of the groups.

Figure 2 shows the effect of the population size N on the density of isolates and on the mean group size for $\alpha = 0$. The results are qualitatively similar to other values of the attractiveness exponent. We note the remarkable unresponsiveness of the density of isolates to changes in N . In fact, although the decrease in N results in a reduction in the rate of events v , given by Equation (13), the effect of a single event on the density of isolates is enhanced. These two effects compensate for each other, resulting in a system size invariance of $\langle n_1(t) \rangle / N$. However, the mean group size decreases with increasing N and converges very rapidly to the infinite system size limit. For $\alpha \leq 1$, this limit is described very well by the analytical approximation for the equilibrium solutions of Equations (6)–(8) that we will derive in the next section.

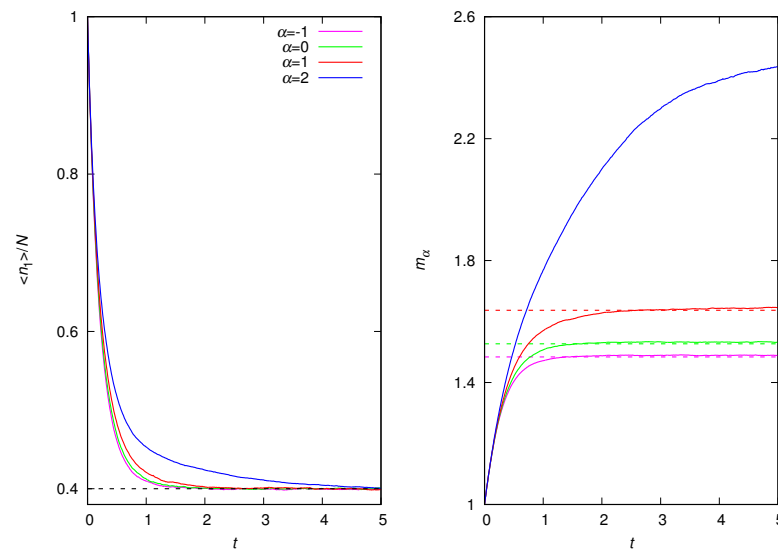


Figure 1. Stochastic simulations of casual groups with attractiveness exponents $\alpha = -1, 0, 1$ and 2 , as indicated. **(Left)** Mean density of isolates as function of time t . The dashed horizontal line is the exact result (11) for the equilibrium regime. **(Right)** Mean group size as function of time t . The dashed horizontal lines are the equilibrium approximate analytical results for $\alpha \leq 1$ presented in Section 4. The other parameters are $\lambda = 1.5$, $\mu = 1$ and $N = 100$.

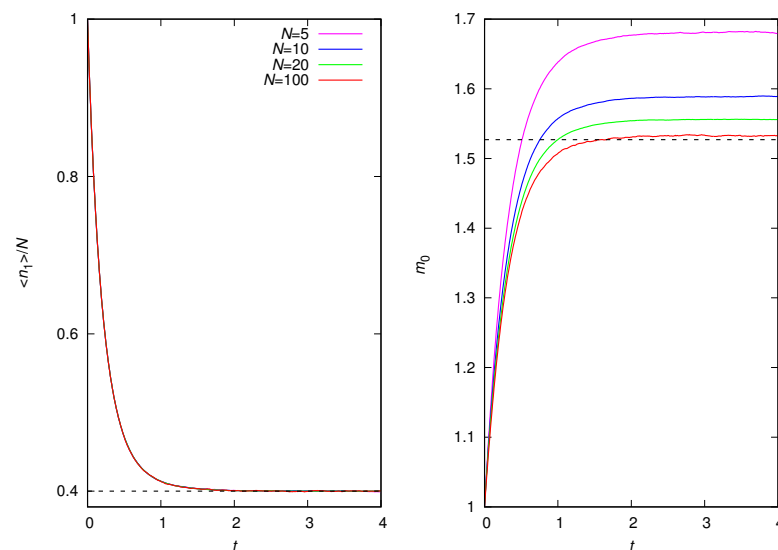


Figure 2. Stochastic simulations of casual groups for $\alpha = 0$ and population sizes $N = 5, 10, 20$ and 100 , as indicated. **(Left)** Mean density of isolates as function of time t . The dashed horizontal line is the exact result (11) for the equilibrium regime. **(Right)** Mean group size as function of time t . The dashed horizontal line is the approximate analytical result for the equilibrium regime presented in Section 4. The other parameters are $\lambda = 1.5$ and $\mu = 1$.

4. Analytical Approximation for the Equilibrium Regime

Here, we derive the controversial approximate analytical results for the expectations $\mathbb{E}[n_i(t)] \equiv \langle n_i(t) \rangle$ that motivated the present contribution [4–6]. In the following, we will focus only on the equilibrium regime, i.e., $d\langle n_i \rangle / dt = 0$ for $i = 1, \dots, N$. To solve the equilibrium equations for $\langle n_i \rangle_{eq}$ with $i > 1$, we make the assumption

$$\mathbb{E}[f(n_1, \dots, n_N)] = f(\langle n_1 \rangle_{eq}, \dots, \langle n_N \rangle_{eq}), \quad (17)$$

where f is an arbitrary rational function. Such a strong assumption is valid if the random variables n_i are self-averaging, i.e., $n_i/N \rightarrow \langle n_i \rangle_{eq}/N$ for all i in the limit $N \rightarrow \infty$. This neglect of fluctuations is the basis of the popular mean-field approximation of statistical physics [25]. With this assumption, we can easily write $\langle n_i \rangle_{eq}$ for $i > 1$ in terms of $\langle n_1 \rangle_{eq}$ and $\langle N_\alpha \rangle_{eq}$,

$$\langle n_i \rangle_{eq} = \frac{[(i-1)!]^{\alpha-1}}{i} \left[\kappa \frac{\langle n_1 \rangle_{eq}}{\langle N_\alpha \rangle_{eq} - 1} \right]^{i-1} (\langle n_1 \rangle_{eq} - 1) \quad (18)$$

where $\langle N_\alpha \rangle_{eq} = \sum_{i=1}^N i^\alpha \langle n_i \rangle_{eq}$, and $\langle n_1 \rangle_{eq}$ is given by Equation (11). These equations must be solved self-consistently: for an arbitrary value of $\langle N_\alpha \rangle_{eq}$, we calculate $\langle n_i \rangle_{eq}$ for $i > 1$, which we then use to update the estimate of $\langle N_\alpha \rangle_{eq}$. The process is repeated until convergence. At this point, we can already see that the assumption in (17) leads to nonphysical results for $\kappa > N - 1$, where $\langle n_1 \rangle_{eq} < 1$ (see Equation (11)), since it implies $\langle n_i \rangle_{eq} < 0$ for $i > 1$. This breakdown of the mean-field approximation is expected because a necessary condition for the self-averaging property to hold is that $\langle n_1 \rangle_{eq} \gg 1$ for large N .

In addition, and more importantly, Equation (18) holds for $\alpha \leq 1$ only. Although for finite N , the self-consistent strategy yields a solution for any value of α , in which $\langle N_\alpha \rangle_{eq}$ scales with N^α , the solution does not satisfy the constraint $\sum_{i=1}^N i \langle n_i \rangle_{eq} = N$ for $\alpha > 1$. The reason is that Equation (18) yields a non-negligible value for $\langle n_N \rangle_{eq}$, resulting in a net flow of individuals to nonphysical group sizes and the consequent violation of the constant-population-size constraint. This effect is negligible for $\alpha \leq 1$ because $\langle n_N \rangle_{eq}$ is vanishingly small (provided that N is not too small), so the flow of individuals to nonphysical regions is inconsequential. We stress that the exact simulations of the group dynamics result in negligible values of $\langle n_N \rangle_{eq}$ for any α , as we will show in Section 5.

A relevant quantity that is usually observed in empirical investigations [1–3] is the mean fraction of groups of size $i = 1, \dots, N$ at equilibrium, defined as

$$p_i = \langle \frac{n_i}{M} \rangle_{eq} \approx \frac{\langle n_i \rangle_{eq}}{\langle M \rangle_{eq}}, \quad (19)$$

where the approximation is justified by the assumption in (17). Of course, p_i can be interpreted as the probability of observing a group of size i . It is interesting that empirical studies typically clump together groups of the same size that are observed on many different occasions (e.g., pedestrians on a sidewalk during Spring mornings in Eugene, Oregon [3]), so they report the total number of observations of groups of a given size. Summing over the different sizes yields the total number of groups observed. Hence, the ratio between the two averages $\langle n_i \rangle_{eq} / \langle M \rangle_{eq}$ is actually the correct measure to describe the empirical results. However, in stochastic simulations, we calculate the ratios n_i/M for each run and then average the results over the many independent runs, so we measure $\langle n_i/M \rangle_{eq}$.

In the following, we present explicit analytical expressions of p_i for $\alpha = 0$, $\alpha = 1$ and the limit $\alpha \rightarrow -\infty$. In addition, we present the numerical solution of Equation (18) obtained with the self-consistent method for general $\alpha \leq 1$.

4.1. Case $\alpha = 0$

In this case, $\langle N_\alpha \rangle_{eq} = \langle M \rangle_{eq}$, and Equation (18) reduces to

$$\langle n_i \rangle_{eq} = \frac{1}{i!} \left[\kappa \frac{\langle n_1 \rangle_{eq}}{\langle M \rangle_{eq} - 1} \right]^{i-1} (\langle n_1 \rangle_{eq} - 1) \quad (20)$$

for $i > 1$, with

$$\begin{aligned}\langle M \rangle_{eq} &= \sum_{i=1}^N \langle n_i \rangle_{eq} \\ &\approx 1 + \frac{1}{\kappa} (\langle M \rangle_{eq} - 1) \left(1 - \frac{1}{\langle n_1 \rangle_{eq}} \right) [\exp(\kappa \frac{\langle n_1 \rangle_{eq}}{\langle M \rangle_{eq} - 1}) - 1],\end{aligned}\quad (21)$$

from which we obtain an explicit expression for the mean number of groups in the equilibrium regime,

$$\langle M \rangle_{eq} \approx \frac{\kappa}{1 + \kappa} \frac{N}{\ln(1 + \kappa)}.\quad (22)$$

In deriving Equation (21), we have assumed that $N \rightarrow \infty$ in order to carry out the sum over the group sizes, whereas, in deriving Equation (22), we have assumed that $\langle n_1 \rangle_{eq}$ is on the order of N , which means that $\kappa \ll N$. As already pointed out, these are the necessary conditions for the validity of the self-averaging property that underlies the mean-field approximation.

Hence,

$$p_i = \frac{1}{e^a - 1} \frac{a^i}{i!},\quad (23)$$

where $a = \ln(1 + \kappa)$, which we identify as the zero-truncated Poisson distribution. Interestingly, this distribution fits a wide variety of data of small groups [3]. The mean and the variance of the group size are

$$m_0 = \sum_{i=1}^N i p_i = (1 + 1/\kappa) \ln(1 + \kappa)\quad (24)$$

and

$$\sigma_0^2 = (1 + 1/\kappa) [\ln(1 + \kappa) + \ln^2(1 + \kappa)] - (1 + 1/\kappa)^2 \ln^2(1 + \kappa).\quad (25)$$

Figure 3 exhibits the comparison between the stochastic simulations and the truncated Poisson distribution (23) for $N = 10$ and $N = 100$. As expected, the mean-field approximation fails to describe the size distribution for $N = 10$, but for $N = 100$, its predictions are indistinguishable from the simulation results. This finding validates the use of that approximation, provided that the number of individuals is not too small.

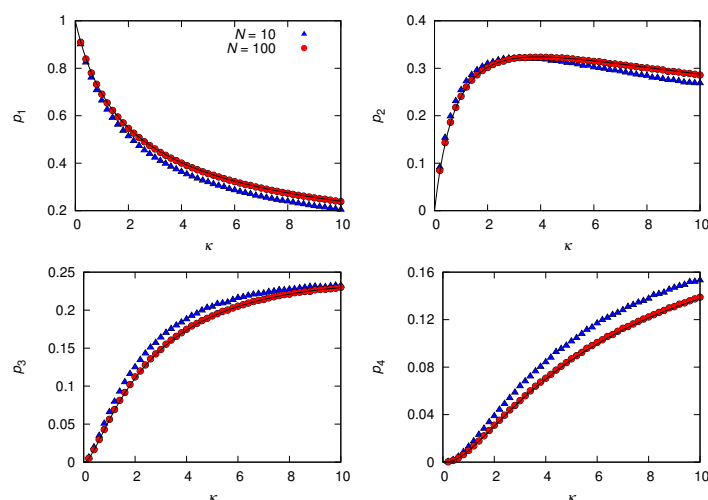


Figure 3. Probability of observing a group of size $i = 1, 2, 3$ and 4 in the equilibrium regime for $\alpha = 0$. The population sizes are $N = 10$ and 100 , as indicated, and the disaggregation rate is $\mu = 1$. The solid curves that perfectly fit the simulation data for $N = 100$ are given by the truncated Poisson distribution (23).

4.2. Case $\alpha = 1$

In this case, $\langle N_\alpha \rangle_{eq} = N$, and Equation (18) reduces to

$$\langle n_i \rangle_{eq} = \frac{1}{i} \left[\kappa \frac{\langle n_1 \rangle_{eq}}{N-1} \right]^{i-1} (\langle n_1 \rangle_{eq} - 1) \quad (26)$$

with $\langle n_1 \rangle_{eq}$ given in Equation (11). As before, assuming that $N \gg 1$ and $\langle n_1 \rangle_{eq} \gg 1$, we obtain an explicit expression for the mean number of groups:

$$\langle M \rangle_{eq} \approx N \frac{1}{\kappa} \ln(1 + \kappa). \quad (27)$$

Thus, the fraction p_i of groups of size $i > 0$ is

$$p_i = \frac{1}{\ln(1 + \kappa)} \frac{1}{i} \left(\frac{\kappa}{1 + \kappa} \right)^i, \quad (28)$$

which we identify as the logarithmic distribution used to model relative species abundance [26]. Hence, the mean group size is

$$m_1 = \frac{\kappa}{\ln(1 + \kappa)} \quad (29)$$

and the variance of the group size is

$$\sigma_1^2 = \frac{-\kappa^2 + \kappa(1 + \kappa) \ln(1 + \kappa)}{\ln^2(1 + \kappa)}. \quad (30)$$

Figure 4 exhibits the comparison between the stochastic simulations and the logarithmic distribution (28) for $N = 10$ and $N = 100$. As before, the results validate the use of the mean-field approximation if the number of individuals is not too small.

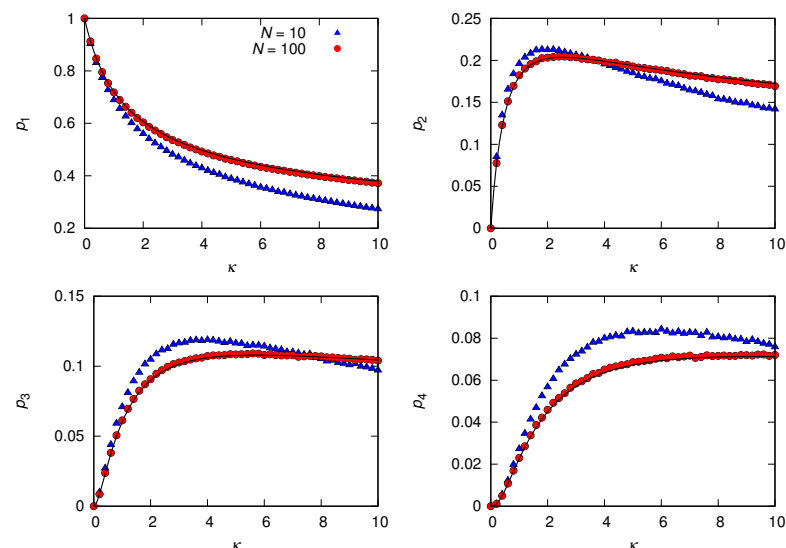


Figure 4. Probability of observing a group of size $i = 1, 2, 3$ and 4 in the equilibrium regime for $\alpha = 1$. The population sizes are $N = 10$ and 100 , as indicated, and the disaggregation rate is $\mu = 1$. The solid curves that perfectly fit the simulation data for $N = 100$ are given by the logarithmic distribution (28).

4.3. The Limit $\alpha \rightarrow -\infty$

In the limit $\alpha \rightarrow -\infty$, we have $\langle N_{-\infty} \rangle_{eq} = \langle n_1 \rangle_{eq}$, $\langle n_2 \rangle_{eq} = \kappa \langle n_1 \rangle_{eq} / 2$ and $\langle n_i \rangle_{eq} = 0$ for $i > 2$. Hence, the fraction of groups of size i is

$$p_1 = \frac{1}{1 + \kappa/2} \quad (31)$$

$$p_2 = \frac{\kappa/2}{1 + \kappa/2} \quad (32)$$

and $p_i = 0$ for $i > 2$. Figure 5 shows the simulation results for $N = 10$ and $N = 100$. We have verified that $p_i = 0$ for $i > 2$ in the simulations. As before, although the mean-field approximation fails for $N = 10$, it yields the exact result for $N = 100$.

The mean group size is

$$m_{-\infty} = \frac{1 + \kappa}{1 + \kappa/2} \quad (33)$$

and the variance of the group size is

$$\sigma_{-\infty}^2 = \frac{\kappa/2}{(1 + \kappa/2)^2}. \quad (34)$$

The limit $\alpha \rightarrow -\infty$ is instructive because we can solve Equations (6)–(8) exactly for any value of $N > 1$ using the fact that $i^\alpha \rightarrow 0$ for $i > 1$. We find

$$\langle n_1(t) \rangle = \langle n_1(0) \rangle e^{-2(1+\kappa)\tau} + \frac{N}{1+\kappa} \left(1 - e^{-2(1+\kappa)\tau} \right) \quad (35)$$

where $\tau = \mu t$. From this equation, we can easily obtain $\langle n_2(t) \rangle = (N - \langle n_1(t) \rangle) / 2$ and $\langle M(t) \rangle = (N + \langle n_1(t) \rangle) / 2$. The reason that Equations (31) and (32) are only approximate and thus fail to fit the data for $N = 10$ in Figure 5 is that, although we can calculate $\langle n_1 \rangle$ and $\langle M \rangle$ exactly for all $N > 1$, we do not know how to calculate $p_1 = \langle n_1 / M \rangle$, which is the quantity measured in the simulations.

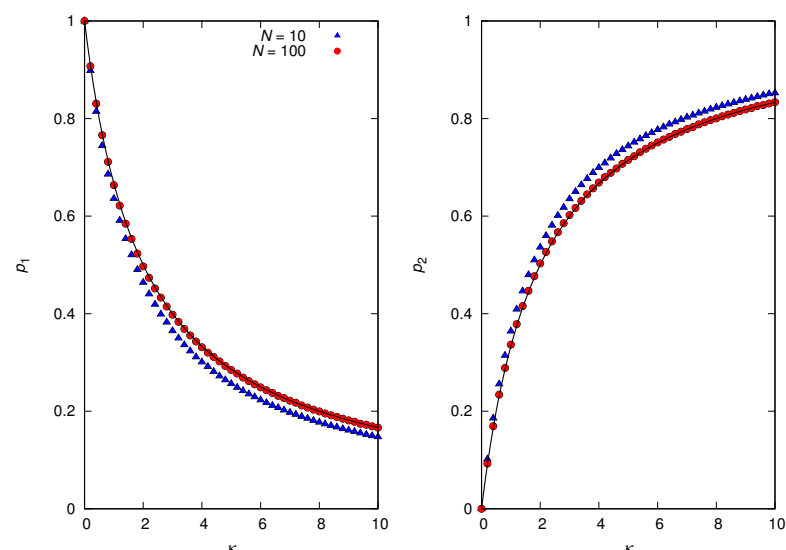


Figure 5. Probability of observing a group of size $i = 1$ and 2 in the equilibrium regime for $\alpha = -10^4$. The population sizes are $N = 10$ and 100 , as indicated, and the disaggregation rate is $\mu = 1$. The solid curves that perfectly fit the simulation data for $N = 100$ are given by Equations (31) and (32).

4.4. General $\alpha \leq 1$

Except for the three cases discussed before, it is not possible to obtain explicit analytical expressions for $\langle n_i \rangle_{eq}$ because we cannot carry out the summation necessary to compute

$\langle N_\alpha \rangle_{eq}$ in a closed form. However, the use of the self-consistent method allows us to easily obtain these quantities numerically. Since we have already established that the mean-field approximation is very accurate, even for $N = 100$, in Figure 6, we present only the approximate theoretical results for m_α and σ_α^2 .

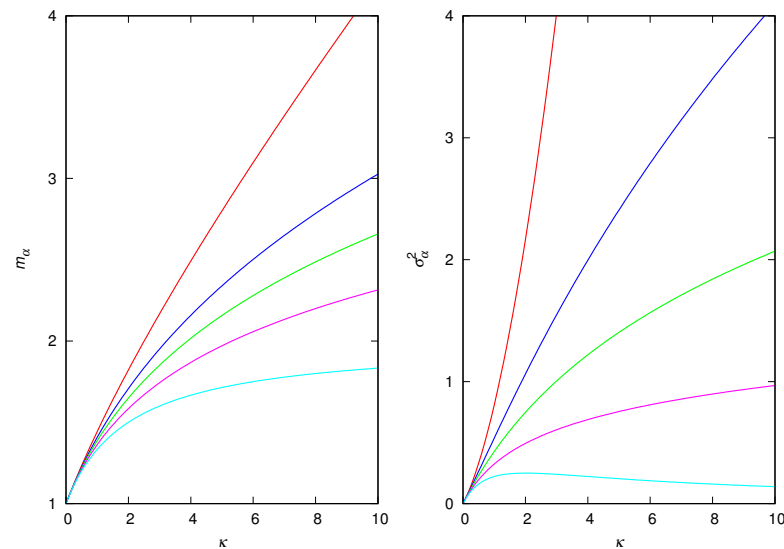


Figure 6. Mean-field approximation for the equilibrium regime for attractiveness exponents (top to bottom) $\alpha = 1, 0.5, 0, -1$ and $\alpha \rightarrow -\infty$. **(Left)** Mean group size m_α . **(Right)** Variance of the group size σ_α^2 .

As expected, decreasing the value of the attractiveness exponent decreases the mean and the variance of the group sizes. We note that these quantities have explicit analytical expressions in the cases $\alpha = 0$ (Equations (24) and (25)), $\alpha = 1$ (Equations (29) and (30)) and $\alpha \rightarrow -\infty$ (Equations (33) and (34)).

5. Equilibrium Regime for $\alpha > 1$

This is by far the most interesting situation because of the complete failure of the mean-field approximation. As already pointed out, the reason is that the solution of Equation (18) violates the fixed-population constraint for $\alpha > 1$. However, we can still obtain some useful analytical information by considering the limit of very large α . In this limit, the system is composed of $\langle n_1 \rangle_{eq}$ isolates and a single group of $N - \langle n_1 \rangle_{eq}$ individuals on average. Hence, the mean number of groups is $\langle M \rangle_{eq} = \langle n_1 \rangle_{eq} + 1$, and we have

$$m_\infty \approx \frac{\langle n_1 \rangle_{eq} + (N - \langle n_1 \rangle_{eq})}{\langle n_1 \rangle_{eq} + 1} \approx 1 + \kappa \quad (36)$$

and

$$\sigma_\infty^2 \approx \frac{\langle n_1 \rangle_{eq} + (N - \langle n_1 \rangle_{eq})^2}{\langle n_1 \rangle_{eq} + 1} - m_\infty^2 \approx N \frac{\kappa^2}{1 + \kappa}, \quad (37)$$

so that the mean group size is finite, but the variance diverges in the limit $N \rightarrow \infty$. The mean size of the large group is

$$N - \langle n_1 \rangle_{eq} = \frac{N\kappa}{1 + \kappa}. \quad (38)$$

Interestingly, Figure 7 shows that this scenario—a single large group coexisting with isolates—describes the case $\alpha = 2$ very well for large N . Figure 8 corroborates this finding by showing that p_i tends to a bimodal distribution characterized by sharp peaks at $i = 1$ and $i = N - n_1$ in the limit of large N . In addition, this figure shows that p_N (or $\langle n_N \rangle_{eq}$)

is negligibly small for $\alpha > 1$, in disagreement with the prediction (17) of the mean-field approximation.

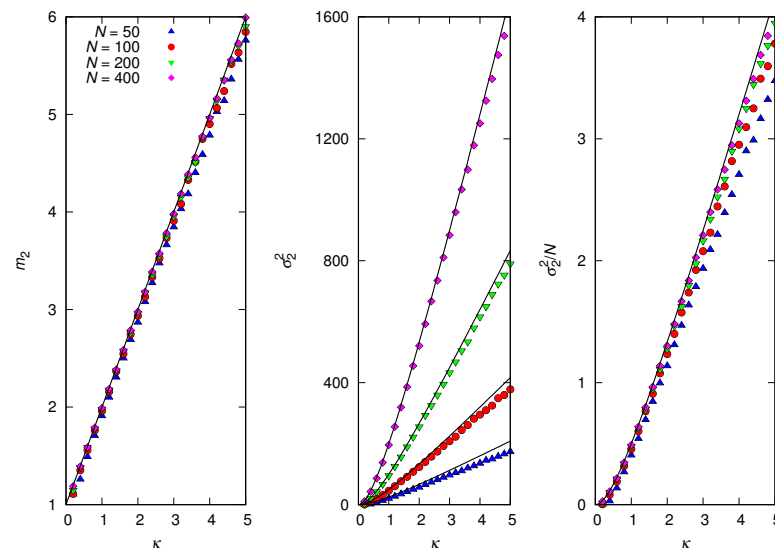


Figure 7. Stochastic simulations in the equilibrium regime for $\alpha = 2$ and population sizes $N = 50, 100, 200$ and 400 , as indicated. (Left) Mean group size m_2 . (Middle) Variance of the group size σ_2^2 . (Right) Scaled variance σ_2^2/N . The solid lines are the predictions of Equations (36) and (37) for $\alpha \rightarrow \infty$. The disaggregation rate is $\mu = 1$.

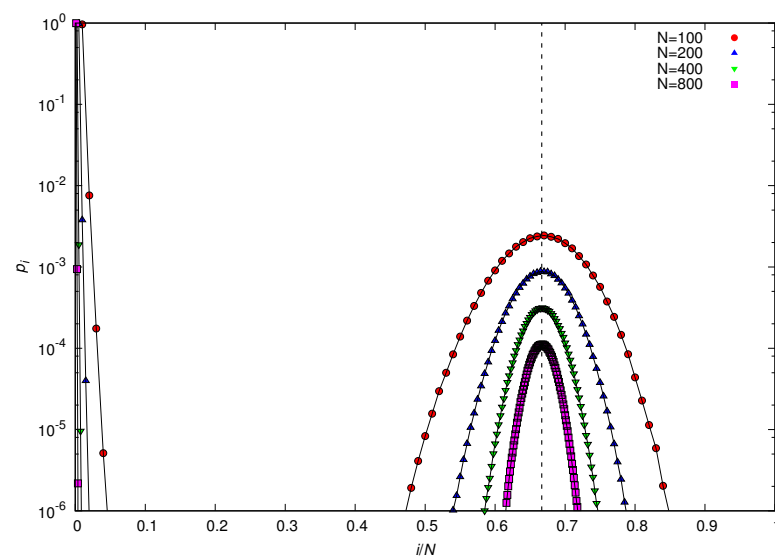


Figure 8. Distribution of group sizes in the equilibrium regime for $\alpha = 2$, $\kappa = 2$, $\mu = 1$ and $N = 100, 200, 400$ and 800 , as indicated. The dashed vertical line indicates the relative size of the large group for $\alpha \rightarrow \infty$, viz., $1 - \langle n_1 \rangle_{eq}/N = 2/3$. The lines connecting the symbols are guides to the eye.

In order to better understand the transition between the equilibrium regime characterized by a finite variance σ_α^2 and the regime where σ_α^2 diverges linearly with N as $N \rightarrow \infty$, in Figure 9, we show the influence of the attractiveness exponent α on m_α and σ_α^2/N . The results indicate the existence of a discontinuous transition between these two regimes that takes place at a critical value $\alpha_c = \alpha_c(\kappa) > 1$. In addition, $\lim_{\alpha \rightarrow \alpha_c^+} m_\alpha < m_\infty$, so the regime of infinite variance but finite α is not perfectly described by the $\alpha \rightarrow \infty$ scenario. In particular, for $\kappa = 2$, we find $\alpha_c \approx 1.105$. This estimate was obtained by considering population sizes up to $N = 12800$ and noticing that, for $\alpha = 1.10$, the variance σ_α^2 tends to a fixed value, whereas, for $\alpha = 1.11$, it increases with N . As κ increases, we find that $\alpha_c \rightarrow 1$.

We note that from the statistical physics perspective, the scaled variance σ_α^2/N is the order parameter of the casual-group model since it is zero for $\alpha < \alpha_c$ and nonzero otherwise.

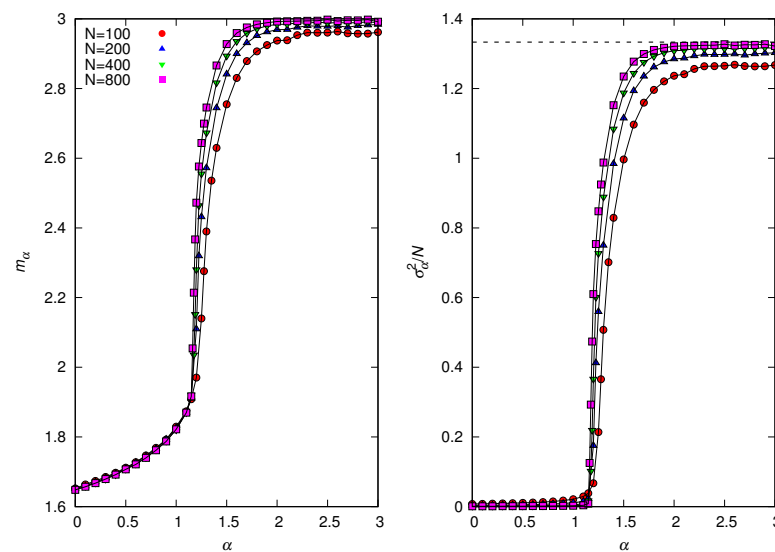


Figure 9. Stochastic simulations in the equilibrium regime for $\kappa = 2$ and population sizes $N = 100, 200, 400$ and 800 , as indicated. **(Left)** Mean group size m_α . **(Right)** Scaled variance σ_α^2/N . The predictions of Equations (36) and (37) for $\alpha \rightarrow \infty$ are $m_\infty = 3$ and $\sigma_\infty^2/N = 4/3$ (dashed horizontal line in the right panel). The disaggregation rate is $\mu = 1$. The lines connecting the symbols are guides to the eye.

Somewhat disappointingly, the acquisition-and-loss process underlying our model does not produce a power-law decay for the probability of finding large groups. In fact, Figure 10 shows that in the vicinity of the transition point α_c , where a scale-free behavior is more likely to be observed [27], the probability of observing large groups vanishes exponentially with increasing group size for $\alpha < \alpha_c$ or exhibits two peaks for $\alpha > \alpha_c$. It remains a challenge to find a simple acquisition-and-loss process that leads to a power-law distribution of group sizes as observed in face-to-face interaction networks [10,11].

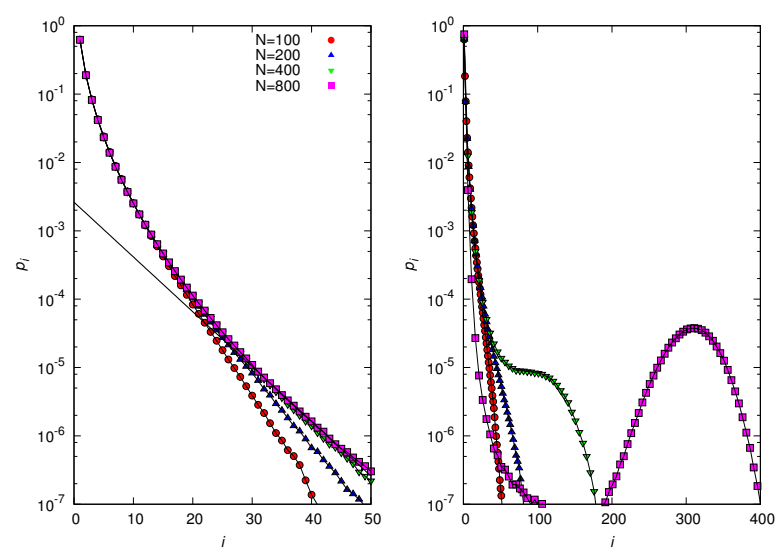


Figure 10. Distribution of group sizes in the equilibrium regime for $\kappa = 2$ and $N = 100, 200, 400$ and 800 , as indicated. **(Left)** $\alpha = 1.10$. The straight line is the fitting $p_i = 0.003 \exp(-0.18i)$ of the data for $N = 800$. **(Right)** $\alpha = 1.15$. The disaggregation rate is $\mu = 1$. The lines connecting the symbols are guides to the eye.

6. Discussion

The distribution of sizes is likely the simplest quantitative information we can derive from the observation of freely forming groups. Of course, if the interrelations of the people present were known a priori, we could most certainly predict the formation and composition of some groups. However, here, we follow an alternative and more fruitful approach that ignores any prior knowledge of the individuals present and attempts to explain the observations using stochastic models characterized by a few parameters that represent people's average tendencies to join and leave a group of a certain size [3,4]. Because the models considered here do not take into account individual idiosyncrasies, we refer to them as null models.

The null models assume that the total number of individuals N is fixed, i.e., that the system is closed, but this assumption is rarely satisfied in field studies of casual groups. For instance, the number of pedestrians on a sidewalk observed on distinct days varies greatly, and it is likely bounded by the city population. However, by taking the limit $N \rightarrow \infty$, the fixed-population constraint becomes inconsequential. Of course, when considering this limit, we must focus only on the ratios of the number of groups, as applied in Equation (19). In fact, this was the approach used in the pioneer paper that introduced the mathematical modeling of casual groups [5]. In any event, our results indicate that even for $N = 100$, the group-size distribution is practically indistinguishable from the distribution derived in the infinite population limit. Regarding the connection between the field studies and the mathematical models, it is assumed that the acquisition-and-loss process takes place at the time of the formation of the groups and that the system is at equilibrium at the moment of the observation. In addition, it is assumed that the observation happens on a time scale that is much faster than the acquisition-and-loss process, so the groups maintain their sizes during the period of observation [3].

Here, we extend previous models of casual groups by assuming that the appeal of a group of size i to isolates is proportional to i^α , where $\alpha \in (-\infty, \infty)$ is the attractiveness exponent. The control of the appeal of groups of different sizes to isolates, which is obtained by tuning the exponent α , allows us to consider some interesting scenarios. For instance, large negative values of α could describe people on the sidewalk walking to a dance party, where a predominance of couples is expected, which cannot be explained by the truncated Poisson distribution (see [28] for an alternative, more complex model of couples and isolates). Large positive values of α result in a bimodal distribution of group sizes, corresponding to a scenario where a single large group coexists with a number of isolates. Interestingly, in both cases, the proportion between the number of individuals in the large group or the number of individuals forming couples and the number of isolates is given by the ratio κ between the rates of aggregation and disaggregation.

Our main result is that the mean-field approximation used to derive the distribution of group sizes in the case that the attractiveness of a group does not depend on its size (i.e., $\alpha = 0$) and in the case that it increases linearly with the group size (i.e., $\alpha = 1$) actually yields the exact result for $N \rightarrow \infty$. This conclusion, which is drawn from the agreement between the exact stochastic simulations of the group dynamics and the mean-field results, dismisses the suspicion of the inadequacy of the mean-field approximation to describe the equilibrium size distribution of casual groups [6]. (Of course, neither Gillespie's algorithm [7,8] nor the computational resources to implement it were available in the 1960s to settle this issue.) In fact, for $\alpha \leq 1$, the mean-field approximation yields very good predictions, even for a relatively small population size (e.g., $N = 100$). However, the approximation fails spectacularly for $\alpha > 1$, since it violates the fixed-population constraint. In this case, Gillespie's stochastic simulation algorithm emerges as the only resource to study the dynamics of casual groups.

In addition, we find that the variation in the attractiveness exponent α produces scenarios where the group sizes are typically small, which is the situation addressed in the literature on casual groups [2], and scenarios where most of the population is confined to a single group. The latter scenario corresponds to the large and stable groups formed by

gregarious animals, whose sizes are determined by a variety of selective pressures, such as defense against predation and foraging success [29]; the cognitive load that constrains the number of individuals with whom it is possible to maintain stable relationships [30]; competence in problem solving [31]; and individual distance preservation [32]. Remarkably, in our model, these distinct scenarios are separated by a discontinuous phase transition that takes place at $\alpha = \alpha_c > 1$, indicating that both types of aggregation behavior can be explained by the same underlying acquisition-and-loss process.

The biological and sociological implication of the success of the null models to produce the empirical distribution of sizes for small groups (i.e., the truncated Poisson distribution) is that prior knowledge of the individuals present in the gathering, as well as individual idiosyncrasies, is not necessary to explain the size distribution of casual groups.

There are at least two research avenues to pursue in order to further improve our understanding of the fleeting clusters of people observed in public gatherings. First, different forms of group attractiveness to isolates can be explored so as to fit the data available from the SocioPatterns collaboration [24], which suggests that the group-size distribution decays as a power law for large group sizes [9–11]. Second, the individual-based model that reproduces the SocioPatterns collaboration data [10,11] can be used to fit the small groups' data available in the seminal works on casual groups [2,3]. If any of these pursuits is successful, one would be able to reproduce all available data on casual groups with a single model.

Funding: This research was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo, grant number 2020/03041-3, and Conselho Nacional de Desenvolvimento Científico e Tecnológico, grant number 305620/2021-5.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Coleman, J.S. *Introduction to Mathematical Sociology*; Free Press Glencoe: London, UK, 1964.
2. Cohen, J.E. *Casual Groups of Monkeys and Men*; Harvard University Press: Cambridge, UK, 1971.
3. Coleman, J.S.; James, J. The Equilibrium Size Distribution of Freely-forming Groups. *Sociometry* **1961**, *24*, 36–45. [\[CrossRef\]](#)
4. White, H. Chance Models of Systems of Casual Groups. *Sociometry* **1962**, *25*, 153–172. [\[CrossRef\]](#)
5. Coleman, J.S. Comment on Harrison White, “Chance Models of Systems of Casual Groups”. *Sociometry* **1962**, *25*, 172–176. [\[CrossRef\]](#)
6. Goodman, L.A. Mathematical Methods for the Study of Systems of Groups. *Am. J. Sociol.* **1964**, *70*, 170–192. [\[CrossRef\]](#)
7. Gillespie, D.T. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comp. Phys.* **1976**, *22*, 403–434. [\[CrossRef\]](#)
8. Gillespie, D.T. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **1977**, *81*, 2340–2361. [\[CrossRef\]](#)
9. Cattuto, C.; Van den Broeck, W.; Barrat, A.; Colizza, V.; Pinton, J.-F.; Vespignani, A. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS ONE* **2010**, *5*, e11596. [\[CrossRef\]](#)
10. Starnini, M.; Baronchelli, A.; Pastor-Satorras, R. Modeling Human Dynamics of Face-to-Face Interaction Networks. *Phys. Rev. Lett.* **2013**, *110*, 168701. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Starnini, M.; Baronchelli, A.; Pastor-Satorras, R. Model reproduces individual, group and collective dynamics of human contact networks. *Soc. Netw.* **2016**, *47*, 130–137. [\[CrossRef\]](#)
12. Lazarsfeld, P.; Berelson, B.; Gaudet, H. *The People's Choice*; Columbia University Press: New York, NY, USA, 1948.
13. Axelrod, R. The Dissemination of Culture: A Model with Local Convergence and Global Polarization. *J. Confl. Res.* **1997**, *41*, 203–226. [\[CrossRef\]](#)
14. Reia, S.M.; Gomes, P.F.; Fontanari, J.F. Comfort-driven mobility produces spatial fragmentation in Axelrod's model. *J. Stat. Mech.* **2020**, *2020*, 033402. [\[CrossRef\]](#)
15. Albert, R.; Barabási A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97. [\[CrossRef\]](#)
16. Girvan, M.; Newman, M.E.J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [\[CrossRef\]](#)
17. Cherifi, H.; Palla, G.; Szymanski, B.K.; Lu, X. On community structure in complex networks: Challenges and opportunities. *Appl. Netw. Sci.* **2019**, *4*, 117. [\[CrossRef\]](#)

18. The Network Data Repository with Interactive Graph Analytics and Visualization. Available online: <https://networkrepository.com> (accessed on 23 April 2023).
19. Sah, P.; Méndez, J.D.; Bansal, S. A multi-species repository of social networks. *Sci. Data* **2019**, *6*, 44. [[CrossRef](#)] [[PubMed](#)]
20. Lancichinetti, A.; Fortunato, S. Community detection algorithms: A comparative analysis. *Phys. Rev. E* **2009**, *80*, 056117. [[CrossRef](#)] [[PubMed](#)]
21. Yang, Z.; Algesheimer, R.; Tessone, C. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Sci. Rep.* **2016**, *6*, 30750. [[CrossRef](#)]
22. Pasquaretta, C.; Levé, M.; Claidière, N.; van de Waal, E.; Whiten, A.; MacIntosh, A.J.J.; Pelé, M.; Bergstrom, M.L.; Borgeaud, C.; Brosnan, S.F.; et al. Social networks in primates: Smart and tolerant species have more efficient networks. *Sci. Rep.* **2014**, *4*, 7600. [[CrossRef](#)] [[PubMed](#)]
23. Fontanari, J.F.; Rodrigues, F.A. Influence of network topology on cooperative problem-solving systems. *Theory Biosci.* **2016**, *135*, 101–110. [[CrossRef](#)]
24. SocioPatterns. Available online: <http://www.sociopatterns.org> (accessed on 23 April 2023).
25. Huang, K. *Statistical Mechanics*; John Wiley & Sons: New York, NY, USA, 1963.
26. Fisher, R.A.; Corbet, A.S.; Williams, C.B. The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. *J. Anim. Ecol.* **1943**, *12*, 42–58. [[CrossRef](#)]
27. West, G. *Scale: The Universal Laws of Life, Growth, and Death in Organisms, Cities, and Companies*; Penguin Press: New York, NY, USA, 2017.
28. Fontanari, J.F. A stochastic model for the influence of social distancing on loneliness. *Phys. A* **2021**, *584*, 126367. [[CrossRef](#)] [[PubMed](#)]
29. Wilson, E. *Sociobiology*; Harvard University Press: Cambridge, MA, USA, 1975.
30. Dunbar, R.I.M. Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* **1992**, *22*, 469–493. [[CrossRef](#)]
31. Fontanari, J.F. Imitative Learning as a Connector of Collective Brains. *PLoS ONE* **2014**, *9*, e110517. [[CrossRef](#)] [[PubMed](#)]
32. Mogilner, A.; Edelstein-Keshet, L.; Bent, L.; Spiros, A. Mutual interactions, potentials, and individual distance in a social aggregation. *J. Math. Biol.* **2003**, *47*, 353–389. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.