Levy geometric graphs

S. Plaszczynski [*], G. Nakamura [†], C. Deroulers, B. Grammaticos, and M. Badoual

Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France
and Université Paris-Cité, IJCLab, 91405 Orsay, France

We present a family of graphs with remarkable properties. They are obtained by connecting the points of a random walk when their distance is smaller than a given scale. Their degree (number of neighbors) does not depend on the graph's size but only on the considered scale. It follows a gamma distribution and thus presents an exponential decay. Levy flights are particular random walks with some power-law increments of infinite variance. When building the geometric graphs from them, we show from dimensional arguments that the number of connected components (clusters) follows an inverse power of the scale. The distribution of the size of their components, properly normalized, is scale invariant, which reflects the self-similar nature of the underlying process. This allows to test if a graph (including nonspatial ones) could possibly result from an underlying Levy process. When the scale increases, these graphs never tend towards a single cluster, the giant component. In other words, while the autocorrelation of the process scales as a power of the distance, they never undergo a phase transition of percolation type. The Levy graphs may find applications in community detection and in the analysis of collective behaviors as in face-to-face interaction networks.

I. INTRODUCTION

Graphs describe *a set of relations* (edges) among some *objects* (vertices) and are thus the fundamental entities for analyzing interactions in complex systems. The celebrated work of Erdös and Rényi [1–3] marks the beginning of graph structure exploration. In this reference model, still often used today to generate null tests, edges are randomly chosen among all possibilities with some given probability p. Many results have been established for these graphs, the most salient feature being that a transition similar to percolation[4] appears beyond some critical connectivity ($p_c = 1/N$, N being the graph size) with a "giant component" containing an extensive number of vertices. The distribution of the number of neighbors in these graphs (called the degree) is a Poisson one and therefore strongly peaked around the mean value pN especially when this one is large.

Graphs embedded in space, i.e., where each vertex has some associated coordinates, are often called "spatial networks." Some adaptation of random graphs to them was proposed [5] by linking nearby points. The resulting graph is now called a *random geometric graph* (RGG). The standard procedure is to first populate randomly N points in the plane and create the edge e_{ji} if the distance between the i and j vertices is below some given cutoff $d(i, j) < R$. The resulting geometric graph is closely related to a pure random one with a connection probability $p = \pi R^2$ (in dimension 2, assuming a unit total surface) and a mean degree

$$\langle k \rangle = pN = \pi R^2 N. \tag{1}$$

RGGs also exhibit a critical transition above which a giant component develops which happens around $\langle k \rangle_c = 4.5$ in two dimensions (2D) [6]. Although there exist some differences between pure random graphs and geometric ones, in particular on the density of triangles, the degree distribution of RGGs is still a Poisson one while many real-world networks are more heavy tailed, going up to power-law (scale-free) distributions [7]. In spatial networks, cost considerations (energetical, economical) tend to restrict the appearance of very large degrees [8], but the degree distributions are still broad.

Several works have focused on ways to obtain a scale-free degree distribution. For RGG, this can be achieved by changing the probability distribution of the points from uniform to a more general form $p(\boldsymbol{x})$ [9], or by changing the space geometry to a hyperbolic one [10]. But the most influential step in that direction is the one by Barabási and Albert [11] who introduced the notion of *growth* (one starts with very few vertices and then adds new ones) and *preferential attachment* (edges are connected depending on the degree of the already present vertices). The success of this approach somewhat shifted the paradigm for graph generation and representation [4] to an iterative process governed by some rules, tightening the links with statistical physics.

Random walks have a long and rich history [12,13] and are of capital importance in statistical physics. By random walk we loosely speak about the repeated sum of the same stochastic processes (steps) and we will restrict ourselves to continuous processes in space. The standard one is based on normally distributed increments (Wiener process) and most walks converge to it since the sum of random variables always converges to a Gaussian thanks to the central limit theorem. This is in fact only valid if the variance of the increment is finite. More generally, the generalized central limit theorem

---
[*]stephane.plaszczynski@ijclab.in2p3.fr
[†]Also at: RIKEN iTHEMS, Wako, Saitama 351-0198, Japan.

[14] states that the sum of any distribution, even with an infinite variance, converges to a stable distribution for which the normal distribution is a particular case.

In what follows we wish to connect the two domains of graph structure exploration and stochastic processes by building a geometric graph from random-walk points. Since power-law interactions are ubiquitous in physics and biology we will put particular emphasis on Levy flights which lead to some remarkable graph properties.

We will first review in Sec. II the fundamentals of Levy flights and the type of geometric graphs produced from them which we shall call Levy geometric graph (LGG), generalizing them to any dimension and discussing the effect of dimensionality. We will then discuss in Sec. III the degree of the graph, making thus a first connection with the random-walk properties. In Sec. IV we study the number and size of the connected components which have some unique properties, and give insights about their structure. Finally, in Sec. V, we shall compare these results to the ones obtained with standard (Gaussian) random walks that will help understand what makes the Levy graphs special. We shall conclude with some possible applications, and defer to more technical Appendixes the computation of the autocorrelation function for a 2D Levy process and of the mean degree of a standard random-walk graph.

II. CONSTRUCTION

A. Levy flight

Mandelbrot [15,16] has introduced the concept of Levy flight (or walk) as a tribute to his teacher's work on stable distributions (for an introduction, see [17]). The method consists first in drawing some radial random number (X) according to a power-law distribution but only above some cutoff value (r_0). Mandelbrot dubbed it the Pareto-Levy distribution. Its cumulative distribution function (also called survival probability) is

$$P(X > r) = \begin{cases} \left(\frac{r_0}{r}\right)^\alpha & \text{for } r \geqslant r_0, \\ 1 & \text{else,} \end{cases} \quad (2)$$

which, by taking the derivative, gives for the probability density function

$$f(r) = \begin{cases} \frac{\alpha}{r_0}\left(\frac{r_0}{r}\right)^{1+\alpha} & \text{for } r \geqslant r_0, \\ 0 & \text{for } r < r_0. \end{cases} \quad (3)$$

An interesting feature of this distribution is that for the Levy index $\alpha < 2$ its variance is infinite, meaning that for samples drawn according to it, the measured standard deviation does not converge with the sample size. From Eq. (2) one derives a straightforward way of drawing numbers according to a Pareto-Levy distribution by first drawing a value u_i from a [0,1] uniform distribution and transforming it according to $r_0 u_i^{-1/\alpha}$. By also drawing an isotropic angle in $[0, 2\pi]$, we obtain the coordinates of a point and build the random walk by accumulating the Euclidean positions [see an example in Fig. 1(a)].

The properties of such a random point process are very unusual to scientists familiar with the convergence properties coming from the central limit theorem, which is not

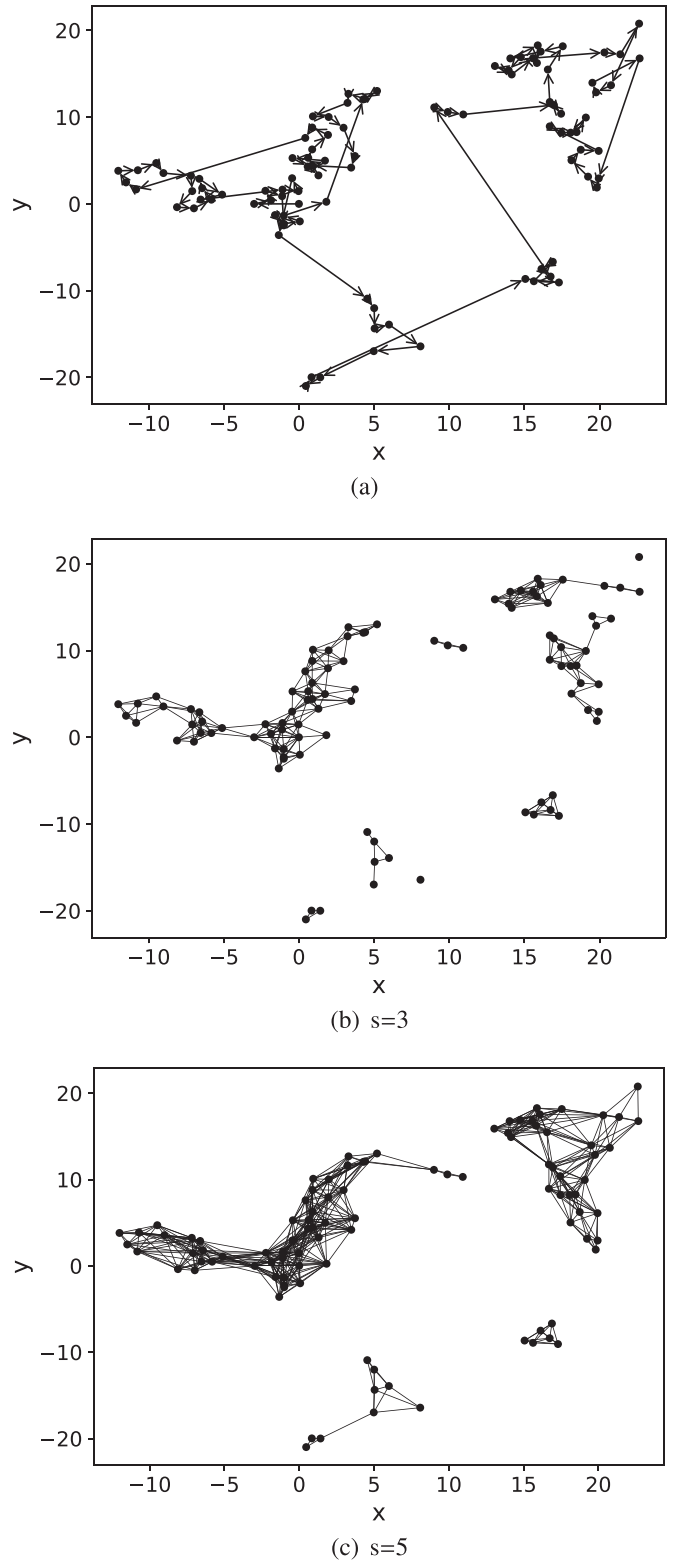FIG. 1. (a) Example of a Levy flight ($\alpha = 1.5$, $r_0 = 1$, $N = 100$) and (b), (c) of two geometric graphs built from it at different scales.

applicable here due to infinite variances. The process is actually *non*homogeneous; there is no mean density as in a Poisson process, or, in the point-process vocabulary [18], a first-order intensity function. However, the process has an isotropic

autocorrelation function (second-order intensity function) defined as the *conditional* probability of finding a point at a distance r from a point of the process. Its computation is explained in detail in Appendix A and leads to

$$f(\boldsymbol{r}) \propto \frac{1}{r^{2-\alpha}} \text{ for } 0 < \alpha < 2 \text{ and } r \gg r_0. \qquad (4)$$

This power-law behavior can be understood considering the asymptotic tail of the Pareto-Levy distribution (3) which is that of a stable distribution [19] with a characteristic function $\phi(k) = e^{-ck^{\alpha}} \simeq 1 - ck^{\alpha}$ in the low k (large r) limit. Fourier integrating it on a space of dimension 2 leads to the result.

By integrating the process on a disk of radius R, one then finds that the mean number of points in it is

$$\bar{N}(< R) \propto \left(\frac{R}{r_0}\right)^{\alpha} \text{ for } 0 < \alpha < 2 \text{ and } R \gg r_0, \qquad (5)$$

exhibiting a fractal dimension in the power law. A process with a power-law autocorrelation function is scale free or more precisely self-similar [20].

We emphasize that these results rely on some approximations that we highlight in Appendix A. In particular, it is sometimes stated that for $\alpha \geqslant 2$ the process becomes Gaussian. Although this is valid for large values of α we show that this transition is progressive. While the power-law description is excellent for α values close to 1, around $\alpha = 2$ the conditional distribution becomes a complicated mixture of power-law and Gaussian functions.

B. Levy geometric graphs (LGG)

The Levy flight is an oriented path. We obtain an undirected graph by applying some scale, i.e., we consider it at some given "resolution." We use the standard geometric graph recipe by connecting points if their Euclidean distance is below some cutoff value R:

$$\|\boldsymbol{X}_i - \boldsymbol{X}_j\| \leqslant R. \qquad (6)$$

What matters here is the relative value between the R cutoff and the minimal step size r_0 of the Pareto-Levy distribution, so that, in what follows, we will only use the *scales* $\equiv \frac{R}{r_0}$ or, equivalently, always work setting r_0 to 1 so that s represents the geometric cutoff.

Increasing the s cutoff, one obtains fewer and fewer clusters which become bigger and bigger as illustrated in Figs. 1(b) and 1(c). Although the resulting graph is a metric one (positions are properties of the vertices) we will only consider their connectivity structure.

For given α exponent and s scale values, we call the resulting graphs the Levy geometric graphs (LGG) and note them $\mathcal{L}_{\alpha}(s)$. The fractal properties are valid for $s \gg 1$ (which will be made more precise in Sec. III) and for $\alpha \leqslant 2$. However for $\alpha < 1$ the mean of the Pareto-Levy distribution diverges and all statistics are governed by rare events leading to very noisy results. So, we shall not consider $\alpha < 1$ values. In what follows, our range of interest for the LGG parameters will be

$$1 \leqslant \alpha \leqslant 2, \qquad (7a)$$

$$s \geqslant 2. \qquad (7b)$$

TABLE I. Return probability as defined in the text measured for Levy graphs with different Levy indices in several dimensions.

d	$\alpha = 1$	$\alpha = 1.5$	$\alpha = 2$
2	0.192 ± 0.002	0.444 ± 0.015	0.709 ± 0.036
3	0.067 ± 0.001	0.141 ± 0.002	0.233 ± 0.002
4	0.031 ± 0.000	0.064 ± 0.001	0.103 ± 0.001
5	0.017 ± 0.001	0.035 ± 0.001	0.055 ± 0.001

C. Dimensionality

Although Levy flights are generally studied in dimension $d = 2$ or 3 we generalize them to any other dimension d by building the walk using 2 for the radius and drawing an isotropic direction, for instance, from a standard d-dimensional normal distribution. The edge assignment is still performed using 6 in the d-dimensional space.

The conditional probability is similar to the 2D case by replacing the exponent 2 in 4 by d. The mean number of points in a ball of radius R [Eq. (5)] is then unchanged up to the normalization factor.

Levy flights may be viewed as a sequence of "local" points followed by some "long" jump. Due to isotropy some new points may "come back" close to some previous ones as in Fig. 1. The probability that this happens, that we call the "return probability," should decrease with dimension, eventually going to 0 as $d \to \infty$ since the path will go to other parts of space.

To be more quantitative, we define a return probability for the Levy process in the following way. Let us first suppose that we have switched "off" the angular part of the process and we only keep the radial steps in an additive way. Then, all pairs of points are separated by a distance of at least $r_0 = 1$. Building a Levy graph that connects points *below* $r_0 = 1$ [$\mathcal{L}_{\alpha}(s = 1)$] just leads to a disconnected set of points where there are as many connected components as points ($N_{\text{clus}} = N$). When switching the angular part "on," some points do come back close to previous ones, sometimes below the $r_0 = 1$ cut, and some clusters start to form ($N_{\text{clus}} < N$). We then propose the following definition for a Levy flight return probability:

$$P_0 = \lim_{N \to \infty} \left(1 - \frac{N_{\text{clus}}}{N}\right), \qquad (8)$$

where N_{clus} is the number of clusters in a $\mathcal{L}_{\alpha}(s = 1)$ of size N.

We estimate those numbers in dimensions 2 to 5 by building 100 $\mathcal{L}_{\alpha}(s = 1)$ graphs ($N = 10^5$), counting each time the number of connected components, and computing the mean and standard deviation of the $(1 - \frac{N_{\text{clus}}}{N})$ values. Results are reported in Table I.

In dimension 2, the return probability is between 19% and 71% depending on the Levy index. If we rescale the $d = 3, 4, 5$ probabilities by $P_0(2)$ we obtain for $P_0(d = 3, 4, 5)/P_0(2)$

$(0.350 \pm 0.006, 0.163 \pm 0.003, 0.089 \pm 0.003) \qquad \alpha = 1,$

$(0.318 \pm 0.011, 0.144 \pm 0.005, 0.078 \pm 0.003) \qquad \alpha = 1.5,$

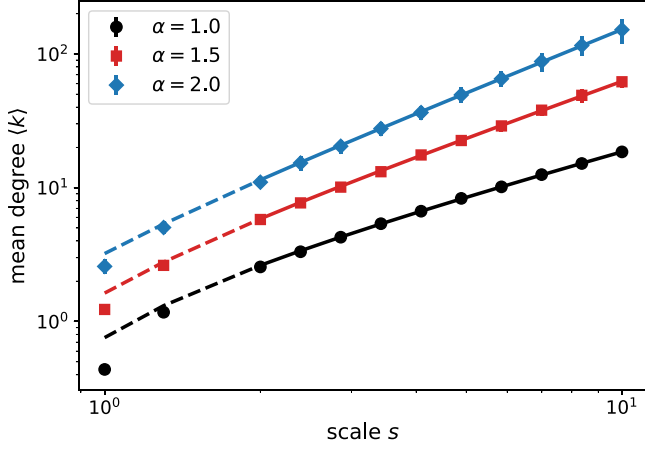$(0.329 \pm 0.017, 0.145 \pm 0.008, 0.078 \pm 0.004) \qquad \alpha = 2,$

FIG. 2. Mean degree measured on dimension-2 Levy graphs varying the scale for several indices values. Full lines represent the Eq. (11) best fits performed in the $s \geqslant 2$ region and the dashed ones their extension to lower values.

which shows in each case a strong effect between dimensions 2 and 3 (about a factor 3), and then milder ones (about a factor 2) when going from dimensions $3 \to 4$ and $4 \to 5$. This effect essentially depends on the space dimension, not on the details of the Levy walk (α). It is worth noticing that these values are similar to the ones obtained for a standard random walk but on a lattice (i.e., a square grid) where the relative probabilities (with respect to dimension 2) to come back to a previous site are [21]

$$P(d)/P(2) = (0.340, 0.193, 0.135). \tag{9}$$

III. DEGREE

We first consider the average degree of the graph. For a geometric graph cut at some distance R, the number of neighbors (degree) at a given vertex is the number of points within a disk of radius R centered on it minus one (the vertex itself). The mean degree is then

$$\langle k \rangle = \bar{N}(< R) - 1. \tag{10}$$

From Eq. (5) we then use the following model for the mean degree:

$$\langle k \rangle(s) = A_D s^{\alpha_D} - 1, \tag{11}$$

where the amplitude A_D and power exponent α_D will be adjusted from the results of simulations.

We measure the mean degree by running 100 $\mathcal{L}_\alpha(s)$ simulations of size $N = 10\,000$ varying the scale and we show the average values with standard deviations for $\alpha = 1, 1.5, 2$ in Fig. 2 together with the best fit to Eq. (11). The agreement is excellent down to $s = 2$ which fixes our lower limit. We have also checked that the power-law model agrees nicely for any Levy index α and in any dimension.

Figure 3 shows the best-fit coefficients in several dimensions. For small values of α, $\alpha_D \simeq \alpha$, but gets smaller when approaching 2. This is to be attributed to the approximations which entered in the derivation of (5) and that are discussed in Appendix A. While α_D is practically independent of the dimension, the amplitude parameter A_D exhibits a strong
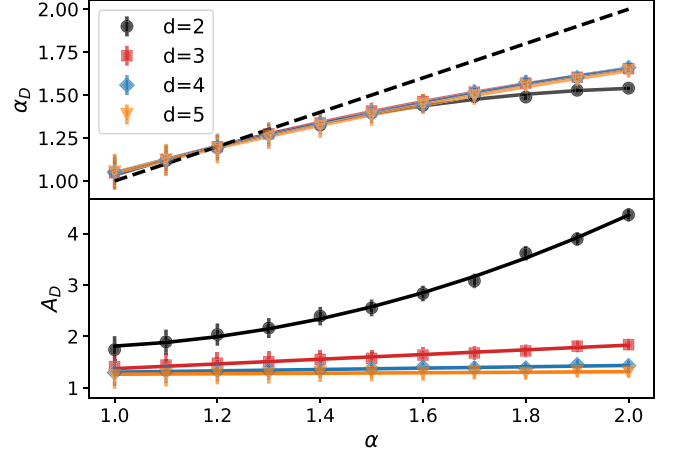

FIG. 3. Best fit parameters values measured on simulations from $\mathcal{L}_\alpha(s)$ mean degree according to the Eq. (11) model in several dimensions d. The points show the measured values and the lines the best quadratic fits (or linear in the case of A_D for $d = 3, 4, 5$). The upper dashed line shows the $\alpha_D = \alpha$ diagonal.

dimension dependence. This is due to the fact that in low dimensions increasing the return probability does increase the mean degree.

In dimension 2, one may use the following approximations:

$$\alpha_D(\alpha) = \alpha - 0.42(\alpha - 1.21)(\alpha - 0.60), \tag{12a}$$

$$A_D(\alpha) = 1.81 + 2.04(\alpha - 1)(\alpha - 0.75), \tag{12b}$$

and we note that the maximal value of α_D is around 1.5.

Finally we emphasize the following:

(i) The mean degree fixes the total number of edges $E = \langle k \rangle \frac{N}{2}$ for undirected graphs. Then, for any $\mathcal{L}_\alpha(s)$ the mean number of edges is known.

(ii) The mean degree of a $\mathcal{L}_\alpha(s)$ is fixed by α and s and is independent of the graph's size N.

The degree distribution has a tail because of points "coming back" to previous ones. We characterized it in Sec. II C by a return probability, that only depends on the space dimension. We have noticed that in our range of parameters Eq. (7) the degree is well described by a Γ distribution

$$P(k) = \frac{k^\beta e^{-k/\theta}}{\Gamma(\beta + 1)\theta^{\beta + 1}}, \tag{13}$$

where $\beta(d)$ depends on the dimension, and we set

$$\theta = \langle k \rangle/(\beta + 1) \tag{14}$$

to ensure the proper mean value since for the Γ distribution $\mathbb{E}[k] = \theta(\beta + 1) = \langle k \rangle$. A fixed value of $\beta = 1.4$ gives good fits for all (α, s) values, as illustrated in Fig. 4. Together with the mean degree formulas $\langle k \rangle(\alpha, s)$ Eqs. (11) and (12), we then obtain an empirical parametrization of the degree distribution for any $\mathcal{L}_\alpha(s)$ (in dimension 2). It shows that for large k the tail decays essentially exponentially.

IV. CONNECTED COMPONENTS

As is clear in Fig. 1, the LGG construction leads to a set of connected components (clusters) which are all simple graphs.
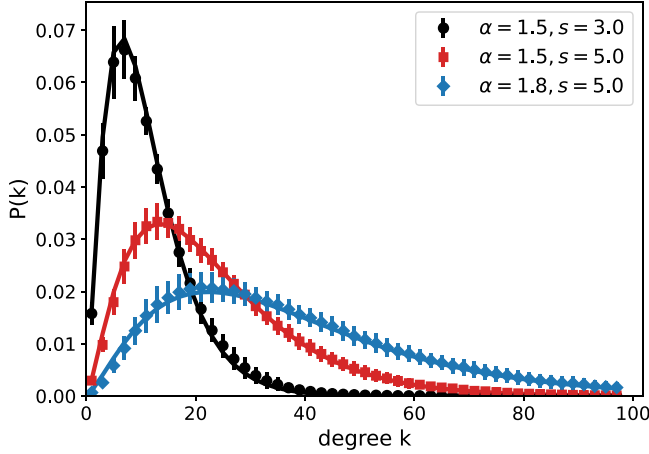
FIG. 4. Parametrization of the degree distribution in dimension 2, for some (α, s) values. The points with error bars show the mean of histograms built from 100 simulations and the line, the analytical formula Eq. (13) with $\beta = 1.4$.

Their number and sizes are random variables which we shall now characterize.

A. Number of clusters

We first look at the number of clusters as a function of the scale for a given Levy index. We measure it for two cases $N = 10^4$ and $N = 10^5$ on simulations ($N_{\text{sim}} = 100$ for each point) by counting the number of connected components. Figure 5 shows the measured cluster fractions for three α values varying the scale. They all follow a power-law function with similar slopes for the two N values in particular when $\alpha \to 1$. As for the mean degree case (Sec. III), the exponent is close to α but here higher by about 25%.

To understand the origin of this scaling we may resort again to the higher-dimensional case where the return probability
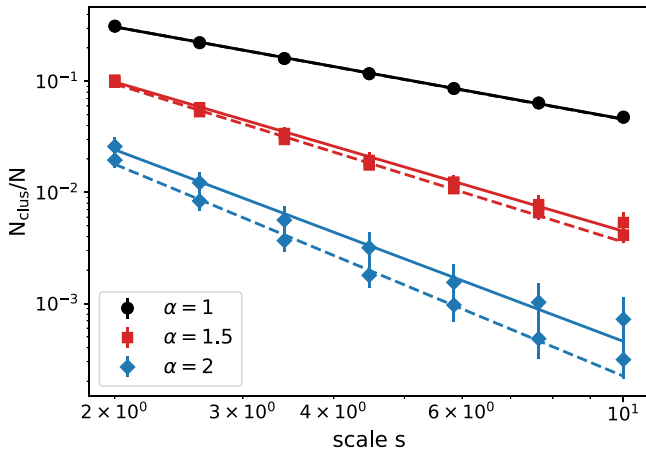

FIG. 5. Measured fraction of clusters (mean and standard deviation over 100 simulations at each point) for three LGGs varying the scale, for two graphs' sizes. Full lines show the the power-law model for $N = 10^4$, and dashed ones for $N = 10^5$. For $\alpha = 1$ both are indistinguishable. The fitted exponents for $\alpha = (1, 1.5, 2)$ are, respectively, $(1.2, 1.9, 2.5)$ for $N = 10^4$ and $(1.2, 2.0, 2.7)$ for $N = 10^5$.
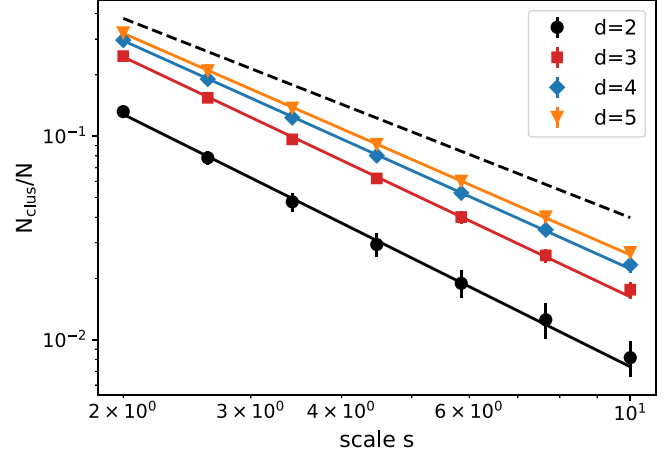

FIG. 6. Measured fraction of clusters of a Levy graph ($\alpha = 1.4$) varying the space dimension $d \in [2, 5]$ ($N = 10^4$). The dashed line shows the asymptotic value $1/s^\alpha$ reached for $d \to \infty$.

may be neglected (Sec. II C). In this case a cluster forms as soon as there is a step larger than the s scale. From Eq. (2) this happens when

$$p(> s) = \frac{1}{s^\alpha}, \tag{15}$$

which shows the power dependency. We show in Fig. 6 how the cluster fraction varies when increasing the dimension. The cluster fraction converges indeed to the Eq. (15) naive expectation following the pattern discussed in Sec. II C (an important change between dimensions 2 and 3 and then some milder ones). The logarithmic slope is unchanged, confirming the fact that the return probability only affects the global normalization.

This also explains why the cluster fraction is mostly independent of N. After a long jump, the probability to have a further one that brings back the walker near a previous point is very small. Clusters are formed in different regions of space so that their number scales about linearly with N.

It is also worth noticing that despite the fact that the process is built from individual steps of infinite variance, the standard deviation on the number of clusters is small. We show in Fig. 7 that the standard deviation on the number of clusters follows $\sigma(N_{\text{clus}}) = b\sqrt{N_{\text{clus}}}$ with $b = 1.4, 2.4, 3$ for, respectively, $\alpha = 1, 1.5, 2$. This is only a factor of around 2 larger than for a Poisson process. This means that for any $\mathcal{L}_\alpha(s)$ graph, the number of clusters in a run of length N is *a priori* known quite precisely.

B. Cluster sizes

We now investigate the cluster sizes, i.e., the number of vertices of each connected component.

For a LGG with N vertices there are N_{clus} clusters of various sizes $N_{i=1,\dots,N_{\text{clus}}}$. Both N_{clus} and N_i's are the realization of random variables subject to the constraint $N = \sum_{i=1}^{N_{\text{clus}}} N_i$. Obviously when there are "fewer" clusters they should be "larger" in order to preserve N. In the following we weight the sizes by the cluster fraction and name it the *normalized*
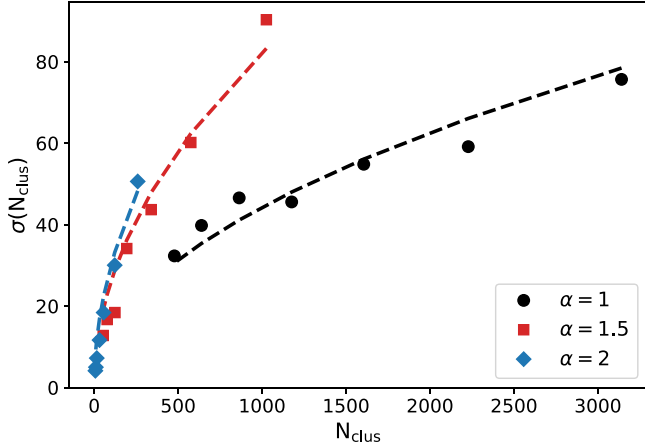
FIG. 7. Standard deviation on the number of clusters as a function of their number for the three runs $\alpha = 1,\ 1.5,$ and 2. The dashed lines show a square-root dependency.

cluster size:

$$n_i = \frac{N_{\text{clus}}}{N} \times N_i \tag{16}$$

and call n the associated random variable.

We show in Fig. 8 the measured survival probability of n for Levy graphs for different indices and scales. The distributions are slightly milder than an exponential one and can be modeled by

$$p(\geqslant n) \propto \exp(-\beta\, n^\gamma) \tag{17}$$

with $\beta \simeq 2$ and $\gamma \simeq 0.4$.

To understand the origin of this shape we consider again the case of a large dimension and show in Fig. 9 the survival probability of n when the dimension increases. The distribu-
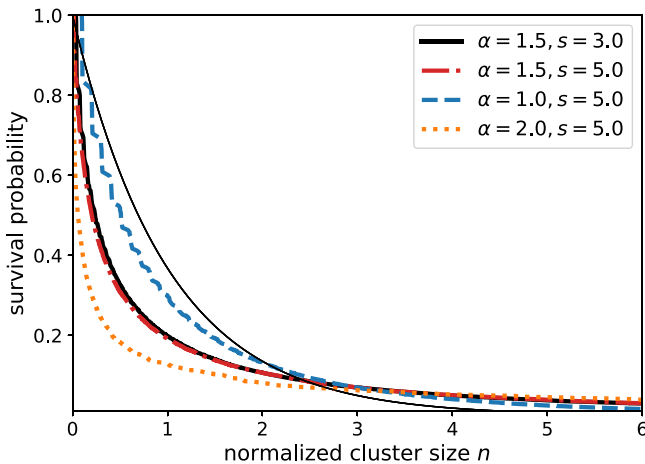

FIG. 8. Survival probability of the normalized cluster size for some LGGs in dimension 2. The first 2 curves (thick solid black and dotted-dashed red) have the same Levy index but different scales. The black one is barely noticeable since both lines superimpose. The following two (dashed blue and dotted orange) show the effect of varying α within the LGG boundaries. The scale used here was 5 but any other value would have given the same result. The thin black line shows the e^{-n} function.
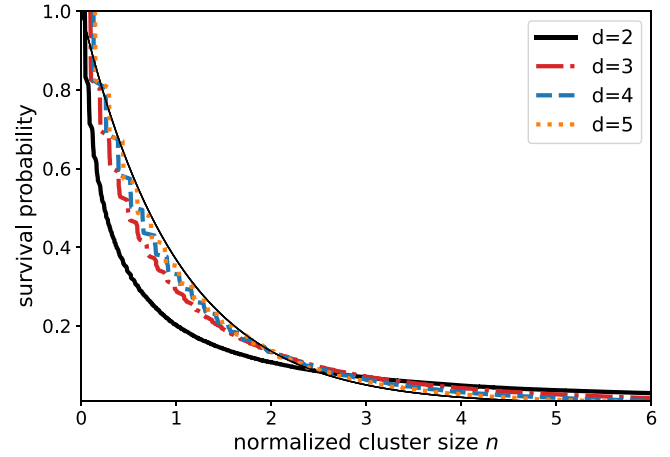

FIG. 9. Survival probability of the (normalized) clusters size when increasing the dimensionality d of the space for $\alpha = 1.5$. They converge to e^{-n} shown as the think black line.

tion becomes closer and closer to an exponential type and seems to converge to e^{-n}. In high dimensions, neglecting the return probability, a cluster of size N_i is formed from several small steps and stops when a jump exceeds the scale s, which happens with probability $p = \frac{1}{s^\alpha}$ [Eq. (2)]. Since the steps are independent, the distribution of the number of points in the cluster is a geometric one:

$$p(N_i) = (1 - p)p^{N_i} \tag{18}$$

$$= \frac{1}{s^\alpha}\left(1 - \frac{1}{s^\alpha}\right)^{N_i}. \tag{19}$$

We have seen that in this space $N_{\text{clus}}/N = 1/s^\alpha$, and by the change of variable $n = \frac{N_{\text{clus}}}{N}N_i$,

$$p(n) = \left(1 - \frac{1}{s^\alpha}\right)^{s^\alpha n}, \tag{20}$$

which, in the region we explore ($s^\alpha \gg 1$), converges indeed to e^{-n}.

But, the most remarkable feature of the Fig. 8 distributions is that they *do not depend on the scale*. As an illustration, we consider the graphs shown in Fig. 1. For $s = 3, 5$ there are, respectively, $N_{\text{clus}} = 9$ and 4 clusters and the normalized sizes are

$$n(s = 3) = \frac{9}{100}(1, 1, 3, 3, 5, 6, 14, 17, 50), \tag{21a}$$

$$n(s = 5) = \frac{4}{100}(6, 9, 32, 53). \tag{21b}$$

If we rank those numbers and plot them on the theoretical curve for $\alpha = 1.5$, we see in Fig. 10 that they are both realizations of the *same* distribution, up to the noise due to the small statistics used for the illustration. Results on a larger statistics is precisely what is shown in Fig. 8.

This statistical invariance comes from the self-similar nature of the Levy flight meaning that the same complexity of the process is contained at any scale. By building the LGG we capture this behavior into the graph. A set of connected components at some given scale is equivalent to any other one built at a different scale. We have thus transferred the fractal geometry of the Levy points to the graph.
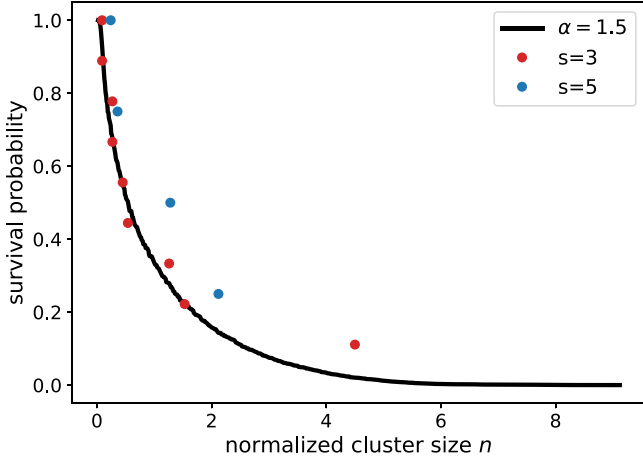
FIG. 10. Normalized cluster size for $\alpha = 1.5$ ($N = 100$). Theoretical curve obtained from simulations in black and realizations observed on Fig. 1 for scales $s = 3$ in red and 5 in blue.

This allows to make a connection to more abstract graphs, i.e., those without a metric (as social networks). From their set of connected components, we can test immediately whether the normalized sizes follow one of the Fig. 8 distributions or not. If not, they are incompatible with a LGG. If yes, we can associate a potential α value, and from the fraction of clusters N_{clus}/N [Eq. (5)], attribute a scale. Further studies then need to be performed to test the topology of the clusters in order to check if the graph could originate from a Levy process. The detailed clusters characterization is outside the scope of this paper and we only illustrate it in the following on the mean degree.

C. Clusters' mean degree

Although the full set of connected components provides an equivalent description of the graph at any scale, a single cluster does not represent the entire graph. Let us call $\langle k \rangle_i$ the mean degree of cluster i:

$$\langle k \rangle_i = \frac{1}{N_i} \sum_{j=1}^{N_i} k_j. \tag{22}$$

The average degree of the graph can then be written

$$\langle k \rangle = \frac{1}{N} \sum_i N_i \langle k \rangle_i = \frac{1}{N_{\text{clus}}} \sum_i n_i \langle k \rangle_i \tag{23}$$

by introducing the normalized cluster sizes n_i [Eq. (16)].

This expression captures the main dependence on the LGG parameters since we have seen that $\langle k \rangle \simeq s^\alpha$ and $N_{\text{clus}} \simeq 1/s^\alpha$. Accordingly, the sum should essentially not depend on s and α. This is shown in Fig. 11 where the distributions of the clusters' mean degree vs their size are similar for different parameters of the LGG.

To understand the global shape, one must remember that the distribution of n is peaked towards low values [Eq. (17)], so we expect many small size components. However, the mean degree of a connected graph is constrained, especially for low sizes. For a cluster of size N_i the smallest degree is achieved with a path ($E_i = N_i - 1$ edges) and the largest one with
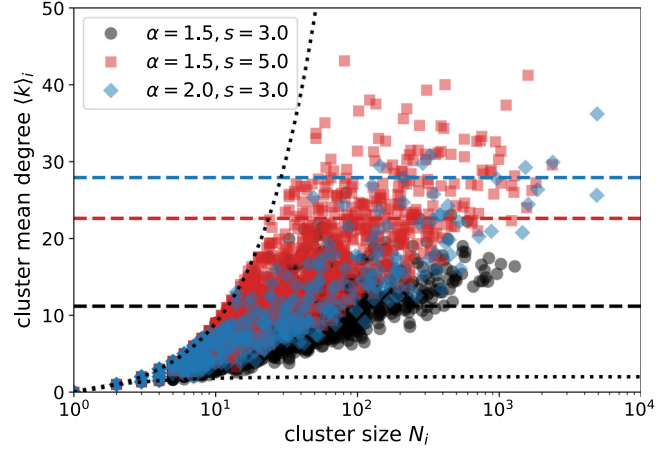

FIG. 11. Mean degree of LGG clusters according to their size. Each point corresponds to one cluster in a $N = 10^5$ simulation. The horizontal dashed lines show the graph's average degree. The dotted lines show the limits discussed in the text [Eq. (24)]. When α or s increases, larger clusters may form for a fixed N size run.

a complete graph [$E_i = \frac{1}{2} N_i (N_i - 1)$]. From $E_i = \langle k \rangle_i \frac{N}{2}$, the bounds on any cluster are therefore

$$2 \frac{N_i - 1}{N_i} \leqslant \langle k \rangle_i \leqslant N_i - 1, \tag{24}$$

corresponding to the gray areas in Fig. 11.

These bounds are very constraining for low size clusters which are the most numerous ones in LGGs. Then, in order to maintain the graph's average degree verifying (23), larger (rare) clusters must have large degrees as observed in Fig. 11. The important point here is that the mean degree is independent of N, so that Fig. 11 is universal. Running with a higher N value, one would (possibly) get a few larger connected components which would add a few points on the right part of the plot, but the main shape would remain unchanged.

Then, each cluster plays a role in obtaining the correct graph's mean degree and a single one cannot be considered as a representation of the whole.

V. RANDOM-WALK GRAPHS

The idea explored in this work is to build a geometric graph on top of a random-walk process. We may then ask what is specific to Levy flights, which are very particular processes with infinite variance steps. We thus compare our results with a geometric graph built on top of a standard random walk (SRW), i.e., with normally distributed increments of variance σ^2.

We first consider the average degree for which we derive an analytical formula in dimension 2 in Appendix B:

$$\langle k \rangle = 2 \sum_{k=1}^{N} \left(1 - \frac{k}{N}\right)\left(1 - e^{-\frac{s^2}{2k}}\right), \tag{25}$$

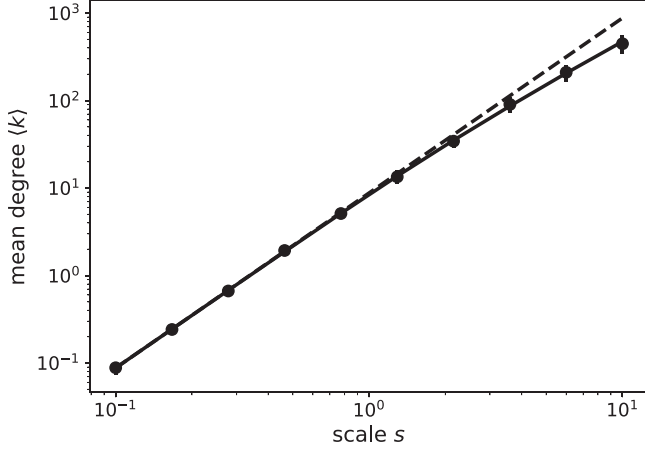where the scale is defined here as $s \equiv \frac{R}{\sigma}$.

FIG. 12. Mean degree for the standard random-walk geometric graph depending on the scale cut $s = \frac{R}{\sigma}$. The points with error bars show the outcome of simulations ($N = 10\,000$, $N_{\text{sim}} = 100$). The full line shows the exact analytical computation Eq. (25), and the dashed one is the quadratic approximation Eq. (26) valid for $s \lesssim 1$.

For $s \lesssim 1$ the argument of the exponential is small, so that

$$\langle k \rangle \simeq s^2 \sum_{k=2}^{N} \frac{1}{k}, \qquad (26)$$

which reveals a quadratic nature but only at low scales. Although formally diverging, the mean degree depends weakly on N in practical cases (the sum being 8.8 for $N = 10^4$ and 13.3 for the $N = 10^6$ case). We confront these calculations to simulations in Fig. 12 showing a perfect agreement.

As for the case of LGG, for which we had $\langle k \rangle \propto s^{\alpha_D}$ with $\alpha_D \lesssim 1.5$ [Eq. (3)], the mean degree for SRWs looks approximately like a power law (with $\alpha_D = 2$). But there is an important difference. While for LGG the formula breaks down at *low* scales [Eq. (2)], for SRW it breaks down at *large* ones [Eq. (12)].

Another similarity comes from the degree distribution. We have checked that for SRWs it is still well described by the Γ distribution (Sec. III). Then, using Eq. (25) we also have an analytical description.

The main difference comes from the clusters. We measure in Fig. 13 the fraction of clusters when increasing the scale, or equivalently the mean degree, and added for reference the RGG case. The SRW graph converges to a single cluster (the giant component) for a connectivity about 10 times larger ($\simeq 50$) than for the RGG. This corresponds to a scale around $s_c = 2$ (see Fig. 12) which is the moment when the mean degree starts to deviate from a pure power law.

For LGG, the power-law behavior stays exact and no giant component ever appears when increasing the scale.[1] This is not only due to the fact that the process is inhomogeneous (which can increase the threshold as in [22] but not suppress the transition), but to the fact that the point density goes to

---

[1]Although technically one could imagine setting the scale to a huge number above the radius of the graph, it cannot be defined *a priori* since the maximal extent of a Levy graph is unpredictable.
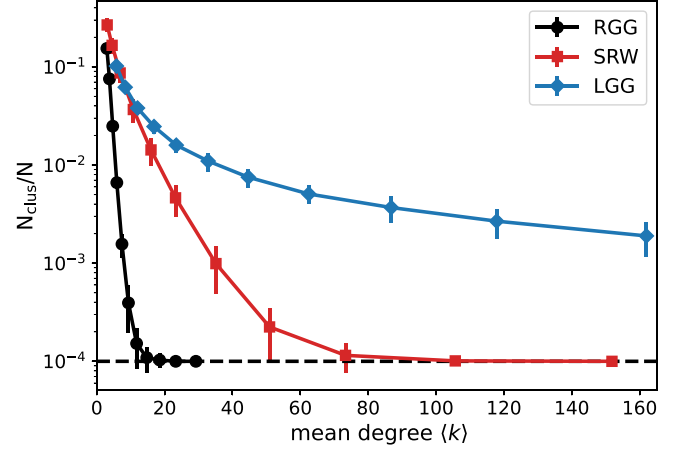

FIG. 13. (a) Fractional number of clusters (for $N = 10^4$) as a function of the mean degree for random geometric graphs (RGG), standard random-walk ones (SRW), and Levy graphs (LGG, $\alpha = 1.5$). The dashed line indicates the $N_{\text{clus}} = 1$ case, i.e., when there is a single giant component.

zero when increasing the geometric cutoff R since $\rho(R) = \frac{\bar{N}(<R)}{\pi R^2} \propto 1/R^{2-\alpha_D}$ with $\alpha_D \lesssim 1.5$ (Sec. III). The set of points is asymptotically *empty*: a randomly placed small volume contains typically no points, which prevents the appearance of the giant component when increasing the radius.

In statistical physics language, the system never undergoes a geometrical phase transition, as in percolation. This type of transition describes the emergence of an ordered phase characterized by giant components: highly connected clusters with sizes of the same order of magnitude as N, i.e., macroscopic structures. At the critical point (or region), though, clusters with various sizes coexist producing large fluctuations in cluster statistics as can be noticed for RGG and SRW in Fig. 13 slightly below the critical connectivity. Traditional random graphs represented here by SRW and RGG can only portray critical behavior in a limited range. In the case of SRW, the typical power-law behavior holds up to scales $s_c \lesssim 2$, indicating that beyond that point a different theory and approximations must be employed to describe the system. In contrast, for LGG the scale invariance remains intact and the same theory can be used, regardless of the scale used to investigate the problem.

VI. CONCLUSION

We have investigated the properties of geometric graphs built on top of random-walk processes and in particular on Levy flights and found the following:

(i) the mean degree is mostly independent of the graph's size;

(ii) it scales as a power law of the geometric cut $\langle k \rangle \propto R^\alpha$, where α is the Levy index (and is equal to 2 for a standard random-walk graph but only for scales below $\sim 2\sigma$);

(iii) the degree follows a Γ distribution and has thus an exponential tail.

These are generic features of all (isotropic) random-walk graphs since from the generalized central limit theorem, any process will either have a finite variance and converge to a

standard (Gaussian) walk, either converge to a stable distribution with Levy-type tails.

We have thus found a simple way to construct a random geometric graph with an exponential tail, i.e., broader than the standard Poisson (Gaussian) one. When considering the connected components (clusters), differences appear between standard random-walk graphs (i.e., with finite variance steps) and Levy-flight graphs (with infinite variance steps). The former show a critical connectivity much larger than for random geometric graphs. But, the latter show *no critical transition at all*. For the Levy graph, a giant component never forms, whatever the scale is.

For Levy graphs the number of clusters scales as an inverse power of the scale. By multiplying it by their size, one obtains a normalized cluster size that is scale invariant, i.e., that does not depend on the geometric cutoff used to build the graph. Thus, the set of clusters *at any scale* is equivalent, which may be viewed as a generalization of the self-similar nature of the Levy flight from points to graphs.

This invariance can allow to make the connection to non-metric graphs by considering only the size of their clusters. If the survival probability falls typically as $e^{-\beta n^\gamma}$ with $\beta \in [2, 3]$ and $\gamma \in [0.3, 1]$, one may associate a potential Levy index, and from the fraction of clusters, a scale. To check further whether a graph could originate from a Levy process or not, one needs to study the structure of its clusters. We have focused on degree distributions but several other topological descriptors exist [7]. We have found, for instance, that the clustering coefficients (that are related to the density of triangles) are large (around 0.7); the average path lengths (shortest number of steps between two vertices) scales as $N^{1/d}$ and is therefore not compatible with a "small-world" network [4]. These two aspects come from the the local nature of the geometric cutoff that favors triangles and forbids the appearance of long shortcuts.

Levy graphs may find application in several areas. On the theoretical side, they reveal an intriguing feature: although they exhibit several power-law dependencies that are characteristic of critical regions [4,23], they actually never experience a transition. Could it be that they are *always* in a critical state? They could then serve as a prototype for studying systems close to a critical point.

Our second finding is that systems without an intrinsic scale but analyzed at a given scale show a very characteristic distribution of their cluster sizes. This may find applications in community detection. Many methods exist to identify communities in a graph but the scale at which to search for them is unclear [24]. Then by running a single algorithm, one can check the cluster characteristics and possibly attribute a Levy index.

Are there some data to which we can confront our model to? To this aim we need to turn to scale-free systems that are common in biology [25], as in the flock of birds [26]. More generally, the analysis and modeling of *collective behaviors* may be an interesting target, as in the self-organization of pedestrian crowds that show some Levy-walk strategies [27]. But, the most direct application could be to the modeling of face-to-face interactions. Some high-quality data that record the time individuals meet in various environments are available [28] and are best analyzed with aggregated graphs [29].

Several important aspects, as the distribution of contact duration, are well described by graphs built on random walks [30]. Biased random walks can also capture the appearance of recurrent communities [31]. It is then natural to explore whether Levy walks may be beneficial to this field since the appearance of communities (clusters) lies at the very heart of Levy graphs.

APPENDIX A: CONDITIONAL PROBABILITY DISTRIBUTION OF A 2D LEVY PROCESS

We detail in this Appendix the computation of the conditional distribution for a Levy process in the plane. We follow closely [33] by adapting it to dimension 2 (since it was performed in dimension 3) enriching the demonstration and quantifying approximations being made.

We start from a point of the process. From (2) the probability distribution of the next displacement in the plane is

$$
f_1(\boldsymbol{r}) = \begin{cases} \dfrac{\alpha}{2\pi} \dfrac{r_0^\alpha}{r^{\alpha+2}} & \text{for } r \geqslant r_0, \\ 0 & \text{otherwise.} \end{cases} \tag{A1}
$$

The process being isotropic, its generating function (Fourier transform) only depends on the mode modulus k. Integrating over the angles

$$
\psi_1(k) = \int f_1(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} d^2\boldsymbol{r} \tag{A2}
$$

$$
= \alpha r_0^\alpha \int_{r_0}^\infty \frac{J_0(kr)}{r^{\alpha+1}} dr, \tag{A3}
$$

where we used [34] [Eq. (7) of Sec. 8.411-7]

$$
\int_0^{2\pi} e^{\pm iz\cos\phi} d\phi = 2\pi J_0(z), \tag{A4}
$$

J_0 being a Bessel function of first type.

Integrating by parts

$$
\psi_1(k) = J_0(kr_0) - kr_0^\alpha \int_{r_0}^\infty \frac{J_1(kr)}{r^\alpha} dr \tag{A5}
$$

using [34] [Eq. (7) of Sec. 6.511-7] $[J_0(kr)]' = -kJ_1(kr)$.

We are interested in the $r \gg r_0$ case so that $kr_0 \ll 1$ and

$$
J_0(kr_0) \simeq 1 - \frac{(kr_0)^2}{4}. \tag{A6}
$$

For $0 < \alpha < 2$ the integral gets most of its contribution from the $r > r_0$ tail so that we can use [34] [Eq. (7) of Sec. 6.561-14]

$$
\int_0^\infty x^\mu J_m(ax) dx = 2^\mu \frac{\Gamma(1/2 + m/2 + \mu/2)}{\Gamma(1/2 + m/2 - \mu/2)} a^{-\mu-1}
$$

$$
\text{for} -m - 1 < \mu < 1/2 \tag{A7}
$$

to obtain

$$
\psi_1(k) \simeq 1 - I_\alpha (kr_0)^\alpha,
$$

$$
\text{with} \quad I_\alpha = \frac{\Gamma(1 - \alpha/2)}{2^\alpha \Gamma(1 + \alpha/2)}. \tag{A8}
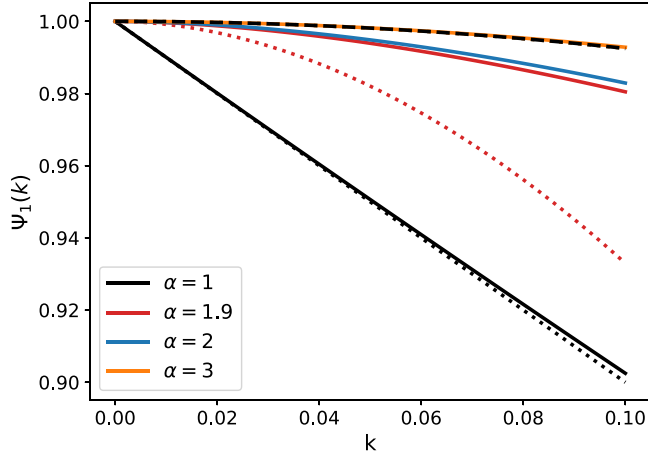$$

FIG. 14. Test of the approximations made in deriving the conditional probability function of a 2D Levy process. The full lines show the exact numerical computation of the single-step characteristic function (A5). Dotted lines show the power-law approximation (A8) used in the $\alpha < 2$ case. The black dashed line shows the quadratic approximation (A14) used for all $\alpha \geqslant 2$ indices. We use $r_0 = 1$ and consider the $kr_0 \ll 1$ region.

One recognizes the asymptotic characteristic function of stable distributions [$\exp(-\sigma^\alpha k^\alpha)$] corresponding to the heavy tail of the Pareto-Levy distribution [19].

The generating function for the nth displacement is the product of the individual functions

$$\psi_n(k) = \psi_1^n(k), \tag{A9}$$

and the probability distribution its inverse Fourier transform

$$f_n(\boldsymbol{r}) = \frac{1}{(2\pi)^2} \int \psi_n(k) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} dk. \tag{A10}$$

Considering *any* number of steps

$$
\begin{aligned}
f(\boldsymbol{r}) &= \sum_n f_n = \frac{1}{(2\pi)^2} \int \sum_n \psi_1^n(k) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} d^2\boldsymbol{k} \\
&= \frac{1}{(2\pi)^2} \int [1 - \psi_1(k)]^{-1} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} d^2\boldsymbol{k} \\
&= \frac{I_\alpha^{-1} r_0^{-\alpha}}{(2\pi)^2} \int k^{-\alpha} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} d^2\boldsymbol{k} \\
&= \frac{I_\alpha^{-1} r_0^{-\alpha}}{2\pi} \int_0^\infty k^{1-\alpha} J_0(kr) dk,
\end{aligned} \tag{A11}
$$

where we use again (A4) when integrating over the angles. From (A7)

$$\int_0^\infty k^{1-\alpha} J_0(kr) dk = K_\alpha r^{\alpha-2} \text{ with } K_\alpha = \frac{\Gamma(1-\alpha/2)}{2^{\alpha-1}\Gamma(\alpha/2)}, \tag{A12}$$

and we finally find that for $\alpha < 2$ and $r \gg r_0$

$$f(\boldsymbol{r}) = \frac{C}{r^{2-\alpha}}, \quad C = \frac{\Gamma(1+\alpha/2)}{\pi\Gamma(\alpha/2)} r_0^{-\alpha}. \tag{A13}$$

For $\alpha \geqslant 2$, the integral (A7) diverges in the $r_0 \to 0$ limit. In fact it now gets most of its contribution from low r values,

i.e., around r_0 where $J_1(kr) \simeq kr/2$. With this crude approximation

$$\psi_1(k) \simeq 1 - \tfrac{3}{4}(kr_0)^2. \tag{A14}$$

This is the leading order of a small Gaussian displacement. Its inverse-Fourier transform is then also a Gaussian and one recovers (roughly) a standard random walk.

We can (and should) question the rather strong simplifications that were made to the (A5) integral in both the $\alpha < 2$ and $\alpha \geqslant 2$ regimes. With $r_0 = 1$, we compare in Fig. 14 the exact value of $\psi_1(k)$ from (A5) computing numerically the integral, to the derived approximations which are Eq. (A8) for $\alpha < 2$ and (A14) for $\alpha \geqslant 2$. The approximation is excellent for $\alpha = 1$ but gets worse when approaching 2. For $\alpha = 2$ the quadratic approximation is not yet reached and becomes satisfactory only around $\alpha = 3$.

APPENDIX B: MEAN DEGREE OF STANDARD RANDOM-WALK GRAPHS

In a standard (Gaussian) random-walk process, the coordinates of the increments follow a normal distribution of variance σ^2, that we note in dimension 2, $x_k, y_k \sim \mathcal{N}(0, \sigma^2)$. The coordinates of the ith point in the walk, as the sum of independent normal variables, then follow $X_i, Y_i \sim \mathcal{N}(0, i\sigma^2)$. Let us focus on a point at index t and compute the distance of any other point at index i to it:

$$r_{ti} = \sqrt{(X_t - X_i)^2 + (Y_t - Y_i)^2}. \tag{B1}$$

Since $X_t - X_i = \sum_{k=1}^t x_k - \sum_{k=1}^i x_k = \sum_{k=i+1}^t x_k$ assuming $i < t$, without loss of generality

$$X_t - X_i \sim \mathcal{N}(0, |t-i|\sigma^2). \tag{B2}$$

The same holds independently for $Y_t - Y_i$, so that (B1) represents the distance between two normally distributed independent variables, each of variance $|t-i|\sigma^2$. It then follows a Rayleigh distribution of cumulative function

$$P_{t,i}(< R) = 1 - e^{-\frac{R^2}{2|t-i|\sigma^2}}. \tag{B3}$$

Let us now consider $\bar{N}_t$ the mean number of points within some distance R of point t. Each point has a probability $P_{t,i}(< R)$ to be in the vicinity of t, so that

$$\bar{N}_t = \sum_{i=1}^N P_{t,i}(< R), \tag{B4}$$

where, for $i = t$, we set $P_{i,i} = 0$ so as to only count neighbors. The mean degree of the geometric graph with an R distance cutoff is obtained by averaging N_t over all the t points:

$$
\begin{aligned}
\langle k \rangle &= \frac{1}{N} \sum_{t=1}^N \bar{N}_t \\
&= \frac{1}{N} \sum_{t=1}^N \sum_{i=1}^N \left(1 - e^{-\frac{s^2}{2|t-i|}}\right),
\end{aligned} \tag{B5}
$$

where we introduce the relevant scale $s \equiv \frac{R}{\sigma}$.

We may simplify the formula by noticing that $P_{t,i}(< R)$ is a circulant matrix symmetric around the $P_{t,t} = 0$ diagonal and

that the double sum represents the sum of all its elements. Then, by counting the elements along the diagonals

$$N\langle k\rangle = 2(N-1)(1-e^{-\frac{s^2}{2}}) + 2(N-2)(1-e^{-\frac{s^2}{4}}) + \cdots \tag{B6}$$

and finally

$$\langle k\rangle = 2\sum_{k=1}^{N}\left(1-\frac{k}{N}\right)\left(1-e^{-\frac{s^2}{2k}}\right). \tag{B7}$$

[1] P. Erdös and A. Rényi, Publicationes Mathematicae **6**, 290 (1959).

[2] P. Erdös and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **6**, 17 (1960).

[3] P. Erdős and A. Rényi, Acta Math. Acad. Sci. Hung. **12**, 261 (1964).

[4] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[5] E. N. Gilbert, J. Soc. Ind. Appl. Math. **9**, 533 (1961).

[6] J. Dall and M. Christensen, Phys. Rev. E **66**, 016121 (2002).

[7] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[8] M. Barthélemy, Phys. Rep. **499**, 1 (2011).

[9] C. Herrmann, M. Barthelemy, and P. Provero, Phys. Rev. E **68**, 026128 (2003).

[10] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, Phys. Rev. E **82**, 036106 (2010).

[11] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[12] N. Van Kampen, in *Stochastic Processes in Physics and Chemistry*, 3rd ed., edited by N. van Kampen (Elsevier, Amsterdam, 2007), pp. 193–218.

[13] B. Hughes, *Random Walks and Random Environments. Volume 1: Random Walks* (Clarendon, Oxford, 1995).

[14] R. Durrett, *Probability: Theory and Examples*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, Cambridge, 2010).

[15] B. Mandelbrot, C. R. (Paris) A **280**, 1551 (1975).

[16] B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1983).

[17] A. V. Chechkin, V. Y. Gonchar, J. Klafter, and R. Metzler, *Fractals, Diffusion, and Relaxation in Disordered Complex Systems* (Wiley, Hoboken, NJ, 2006), pp. 439–496.

[18] D. Cox and V. Isham, *Point Processes*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability (Taylor & Francis, London, 1980).

[19] A. V. Chechkin, R. Metzler, J. Klafter, and V. Y. Gonchar, *Introduction to the Theory of Lévy Flights* (Wiley, Weinheim, Germany, 2008), pp. 129–162.

[20] M. E. J. Newman, Contemp. Phys. **46**, 323 (2005).

[21] E. W. Montroll, J. Soc. Ind. Appl. Math. **4**, 241 (1956).

[22] P. Wang and M. C. González, Philos. Trans. R. Soc. A **367**, 3321 (2009).

[23] H. E. Stanley, Rev. Mod. Phys. **71**, S358 (1999).

[24] S. Fortunato, Phys. Rep. **486**, 75 (2010).

[25] T. Mora and W. Bialek, J. Stat. Phys. **144**, 268 (2011).

[26] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, Proc. Natl. Acad. Sci. USA **107**, 11865 (2010).

[27] H. Murakami, C. Feliciani, and K. Nishinari, J. R. Soc. Interface. **16**, 20180939 (2019).

[28] www.sociopatterns.org.

[29] P. Holme and J. Saramäki, Phys. Rep. **519**, 97 (2012).

[30] M. Starnini, A. Baronchelli, and R. Pastor-Satorras, Phys. Rev. Lett. **110**, 168701 (2013).

[31] M. A. R. Flores and F. Papadopoulos, Phys. Rev. Lett. **121**, 258301 (2018).

[32] https://graph-tool.skewed.de

[33] P. J. E. Peebles, *The large-scale Structure of the Universe* (Princeton University Press, 1980), Chap. III.62, p. 435.

[34] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 2007).