



HAL
open science

Cliques in high-dimensional random geometric graphs

Konstantin Avrachenkov, Andrei V Bobu

► **To cite this version:**

Konstantin Avrachenkov, Andrei V Bobu. Cliques in high-dimensional random geometric graphs. Applied Network Science, 2020, 5, 10.1007/s41109-020-00335-6 . hal-03023124

HAL Id: hal-03023124

<https://inria.hal.science/hal-03023124>

Submitted on 25 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH

Open Access



Cliques in high-dimensional random geometric graphs

Konstantin E. Avrachenkov and Andrei V. Bobu*

*Correspondence:
a.v.bobu@gmail.com
INRIA, 2004, route des
Lucioles, 06902 Valbonne,
France

Abstract

Random geometric graphs have become now a popular object of research. Defined rather simply, these graphs describe real networks much better than classical Erdős–Rényi graphs due to their ability to produce tightly connected communities. The n vertices of a random geometric graph are points in d -dimensional Euclidean space, and two vertices are adjacent if they are close to each other. Many properties of these graphs have been revealed in the case when d is fixed. However, the case of growing dimension d is practically unexplored. This regime corresponds to a real-life situation when one has a data set of n observations with a significant number of features, a quite common case in data science today. In this paper, we study the clique structure of random geometric graphs when $n \rightarrow \infty$, and $d \rightarrow \infty$, and average vertex degree grows significantly slower than n . We show that under these conditions, random geometric graphs do not contain cliques of size 4 a. s. if only $d \gg \log^{1+\epsilon} n$. As for the cliques of size 3, we present new bounds on the expected number of triangles in the case $\log^2 n \ll d \ll \log^3 n$ that improve previously known results. In addition, we provide new numerical results showing that the underlying geometry can be detected using the number of triangles even for small n .

Keywords: Random geometric graphs, High dimension, Clique number, Triangles

Introduction

Graphs are the most natural way to model many real-world networks. The links between nodes in a network are often occasional, that is why deterministic graphs are often irrelevant for network modeling. Random graphs, in which the link presence is a random variable, appear in the 20 th century with the simplest model proposed by Erdős and Rényi (see Erdős and Rényi 1960; Bollobás 2001; Alon and Spencer 2004). In this model, the edges between vertices appear independently with equal probability. However, this model fails to describe some important properties of many real networks, such as the community structure. Many models have been proposed to overcome this and other problems: the Barabási–Albert model (see Albert and Barabási 2002, 1999; Fronczak et al. 2003), random geometric graphs (see Gilbert 1961; Penrose 2003), hyperbolic geometric graphs (see Barthélemy 2011; Krioukov et al. 2010). Perhaps random geometric graphs are the simplest natural model where the edge appearance depends only on the Euclidean distance between given

nodes. These graphs resemble real social, technological and biological networks in many aspects. Also, they might be useful in statistics and machine learning tasks: the correlations between observations in a data set can be represented by links between corresponding close vertices. Moreover, analyzing the graph in this situation can help to determine the presence of the underlying geometry and, therefore, the possibility of embedding into some geometric space. A reader can find these and other applications in Preciado and Jadbabaie (2009), Haenggi et al. (2009), Pottie and Kaiser (2000), Nekovee (2007), Xiao and Yeh (2011), Higham et al. (2008), Arias-Castro et al. (2015).

Let us define the model of the graphs that interest us and introduce a notation system. We follow the article Devroye et al. (2011) and define a random geometric graph $G(n, p, d)$ as follows. Let X_1, \dots, X_n be independent random vectors uniformly distributed on the $(d - 1)$ -dimensional sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$. Two distinct vertices $i \in [n]$ and $j \in [n]$ are adjacent if and only if $\langle X_i, X_j \rangle \geq t_{p,d}$; here $t_{p,d}$ is defined in such a manner that $\mathbb{P}(\langle X_i, X_j \rangle \geq t_{p,d}) = p$.

As for fixed d , many properties of random geometric graphs (connectivity, large components, number of small subgraphs) have been revealed at the end of the 20 th century. We refer to Penrose (2003) for an intensive study of random geometric graphs; among other papers on this topic we can highlight papers Arias-Castro et al. (2015), Penrose (1999), Appel and Russo (2002), McDiarmid (2003), Müller (2008), McDiarmid and Müller (2011). However, the lack of results regarding the high-dimensional case $d \rightarrow \infty$ looks quite surprising. This becomes even more remarkable with the growing number of features in data sets and, hence, possible application of high-dimensional random geometric graphs in machine learning problems.

As far as we are concerned, the first paper treating the case $d \rightarrow \infty$ is the article of Devroye et al. (2011). The authors studied the clique number in the asymptotic case when $n \rightarrow \infty$, $d \gg \log n$, and p is fixed (hereinafter $f(n) \gg g(n)$ (or $f(n) \ll g(n)$) means $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = \infty$ (respectively, $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0$)). As it was proven in Devroye et al. (2011), in this *dense regime*, the clique number of the random geometric graph $G(n, p, d)$ is close to that of the Erdős–Rényi graph $G(n, p)$.

Another extremely important paper is the work of Bubeck et al. (2016). The authors considered the *thermodynamic regime* when a node has a constant number of neighbours on average; their paper focuses mainly on the difference between random geometric and Erdős–Rényi graphs. They have obtained a ‘negative’ result: for $d \ll \log^3 n$, the graph $G(n, c/n, d)$ (here c is a constant) is ‘different’ from the Erdős–Rényi graph $G(n, c/n)$ in the sense that the total variation distance between two random models converges to 1 as $n \rightarrow \infty$. Also, the authors made a conjecture about a ‘positive’ result: for $d \gg \log^3 n$, the graphs $G(n, p, d)$ and $G(n, p)$ are close. In order to obtain the ‘negative’ result, they proved that in the case $d \ll \log^3 n$ the average number of triangles in $G(n, c/n, d)$ grows at least as a polylogarithmic function of n , which is different from the expected number of triangles in $G(n, c/n)$. The difference between these regimes seems quite interesting, and that is why we concentrate on the case when d is a power of $\log n$.

Our main interest is the investigation of cliques in the asymptotic case when $n \rightarrow \infty$ and $d \gg \log n$. Since the dense regime (when p is fixed) is well studied in Devroye et al. (2011), we will focus on the *sparse regime* when $p = p(n) = o(1)$ as $n \rightarrow \infty$. Also, we

assume that $p(n)$ does not go to 0 ‘extremely fast’ (for instance, as e^{-n}). This, of course, includes the regime considered in Bubeck et al. (2016).

The main contribution of the present paper consists of three results in the sparse regime. The first one, presented in the next section in Theorem 4, states that almost surely the clique number of $G(n, p, d)$ is at most 3 under the condition $d \gg \log^{1+\epsilon} n$. The second one shows that the expected number of triangles grows as $\binom{n}{3} p^3$ in the case $d \gg \log^3 n$. This result is given in Theorem 5. Finally, in Theorem 7 we will present new lower and upper bounds on the expected number of triangles in the case $\log^2 n \ll d \ll \log^3 n$. This lower bound improves the result of Bubeck et al. (2016) since it grows faster than any polylogarithmic function (recall that the lower bound from Bubeck et al. (2016) is polylogarithmic in n).

The latter result clearly shows that for $d \ll \log^3 n$, random geometric graphs have a greater tendency to form clusters than it was discovered in the paper (Bubeck et al. 2016). We will present some numerical results showing that this is true even for relatively small n . Hence, the number of triangles can be used as a tool to determine if a given graph has some hidden geometry.

This paper is an extended version of the Complex Networks conference paper Avrachenkov and Bobu (2019). Compared to the short paper, we present here absolutely new numerical results and more specifically discuss the application of the present work to real-life tasks of machine learning. Also, we have improved Theorems 4 and 5 and give full proofs of the results mentioned in Avrachenkov and Bobu (2019).

Related works

Let us describe the related works (Devroye et al. 2011) and (Bubeck et al. 2016) in more detail. We start with the results of Devroye et al. (2011). Although this paper is devoted to the dense regime, the following two theorems do not require the condition $p = \text{const}$ and turn out to be useful in our situation. Let us denote by $N_k(n, p, d)$ the number of cliques of size k in $G(n, p, d)$. The results below establish lower and upper bounds on $\mathbb{E}[N_k(n, p, d)]$.

Theorem 1 (Devroye et al. 2011) *Introduce*

$$\tilde{p} = \tilde{p}(p) = 1 - \Phi(2t_{p,d}\sqrt{d} + 1).$$

Let $\delta_n \in (0, 2/3]$, and fix $k \geq 3$. Assume

$$d > \frac{8(k + 1)^2 \log \frac{1}{\tilde{p}}}{\delta_n^2} \left(k \log \frac{4}{\tilde{p}} + \log \frac{k - 1}{2} \right).$$

Define $\alpha > 0$ as

$$\alpha^2 = 1 + \sqrt{\frac{8k}{d} \log \frac{4}{\tilde{p}}}.$$

Then

$$\mathbb{E}[N_k(n, p, d)] \geq \frac{4}{5} \binom{n}{k} \left(1 - \tilde{\Phi}_k(d, p)\right) \binom{k}{2},$$

where $\tilde{\Phi}_k(d, p) = \Phi\left(\frac{\alpha t_{p,d} \sqrt{d} + \delta_n}{\sqrt{1 - \frac{2(k+1)^2 \log(1/\hat{p})}{d}}}\right)$. Here $\Phi(\cdot)$ denotes the CDF of the standard normal distribution.

Theorem 2 (Devroye et al. 2011) *Let $k \geq 2$ be a positive integer, let $\delta_n > 0$, and define*

$$\hat{p} = \hat{p}(p) = 1 - \Phi(t_{p,d} \sqrt{d}).$$

Assume

$$d \geq \frac{8(k+1)^2 \log \frac{1}{\hat{p}}}{\delta_n^2} \left(k \log \frac{4}{\hat{p}} + \log \frac{k-1}{2}\right).$$

Furthermore, for $p < 1/2$, define $\beta = 2\sqrt{\log(4/\hat{p})}$ and for $\beta\sqrt{k/d} < 1$, let $\alpha = \sqrt{1 - \beta\sqrt{k/d}}$. Then for any $0 < \delta_n < \alpha t_{p,d} \sqrt{d}$, we have

$$\mathbb{E}[N_k(n, p, d)] \leq e^{1/\sqrt{2}} \binom{n}{k} \left(1 - \Phi(\alpha t_{p,d} \sqrt{d} - \delta_n)\right) \binom{k}{2}.$$

The following result of Bubeck et al. (2016) gives a lower bound on the expected number of triangles and significantly improves Theorem 1 if only $d \ll \log^3 n$.

Theorem 3 (Bubeck et al. 2016) *There exists a universal constant $C > 0$ such that whenever $p < 1/4$ we have that*

$$\mathbb{E}[N_3(n, p, d)] \geq p^3 \binom{n}{3} \left(1 + C \frac{(\log \frac{1}{p})^{3/2}}{\sqrt{d}}\right).$$

Note that if $p = \theta(n)/n$ with some function $\theta(n) \ll n^{1-\epsilon}$ and $d \ll \log^3 n$, the expected number of triangles grows as a polylogarithmic function. This is totally different from the Erdős–Rényi graph $G(n, \theta(n)/n)$ where the average number of triangles grows as $\theta^3(n)$ with $n \rightarrow \infty$. This result will be further improved by our Theorem 7.

Results

Clique number in the sparse regime

Theorems 1 and 2 allow us to say that for constant p and $d \gg \log^7 n$, the clique number of a random geometric graph grows similarly to that of an Erdős–Rényi graph, which is $2 \log_{1/p} n - 2 \log_{1/p} \log_{1/p} n + O(1)$. We will show that in the sparse regime, under some conditions on p and d , there is no clique of size 4 in $G(n, p, d)$ a. s. Let us also remark that the condition $d \gg \log^7 n$ is necessary only for the dense regime, as Theorems 1 and 2 do not impose any restrictions on d .

To apply Theorems 1 and 2, we need a lemma that establishes the growth rate of $t_{p,d}$, which is crucial for asymptotic analysis in the sparse regime. Since p is the normalized surface area of a spherical cap of angle $\arccos t_{p,d}$ (an example for a circle is given in Fig. 1 below), we learn from convex geometry that (see Brieden et al. 2001):

$$\frac{1}{6t_{p,d}\sqrt{d}}(1 - t_{p,d}^2)^{\frac{d-1}{2}} \leq p \leq \frac{1}{2t_{p,d}\sqrt{d}}(1 - t_{p,d}^2)^{\frac{d-1}{2}}. \tag{1}$$

The following more explicit bound on $t_{p,d}$ has been derived in Bubeck et al. (2016).

Lemma 1 *There exists a universal constant $C > 0$ such that for all $0 < p \leq 1/2$, we have that*

$$\min \left(C^{-1}(1/2 - p)\sqrt{\frac{\log(1/p)}{d}}; \frac{1}{2} \right) \leq t_{p,d} \leq C\sqrt{\frac{\log(1/p)}{d}}.$$

Before we present our main result on the clique number, let us prove a lemma that will be useful later.

Lemma 2 *Let $d \geq \log^{1+\epsilon} n$ for some fixed $\epsilon > 0$ and $p \geq 1/n^m$ for some fixed m . Then for any $\gamma > 0$, there exists n_0 such that for $n \geq n_0$*

$$\frac{e^{-t_{p,d}^2 d/2}}{t_{p,d}\sqrt{d}} \leq pn^\gamma.$$

Proof First of all, let us notice that $t_{p,d} \geq 0$, that is why

$$e^{-t_{p,d}^2 d/2} = e^{-t_{p,d}^2 (d-1)/2} e^{-t_{p,d}^2 /2} \leq e^{-t_{p,d}^2 (d-1)/2},$$

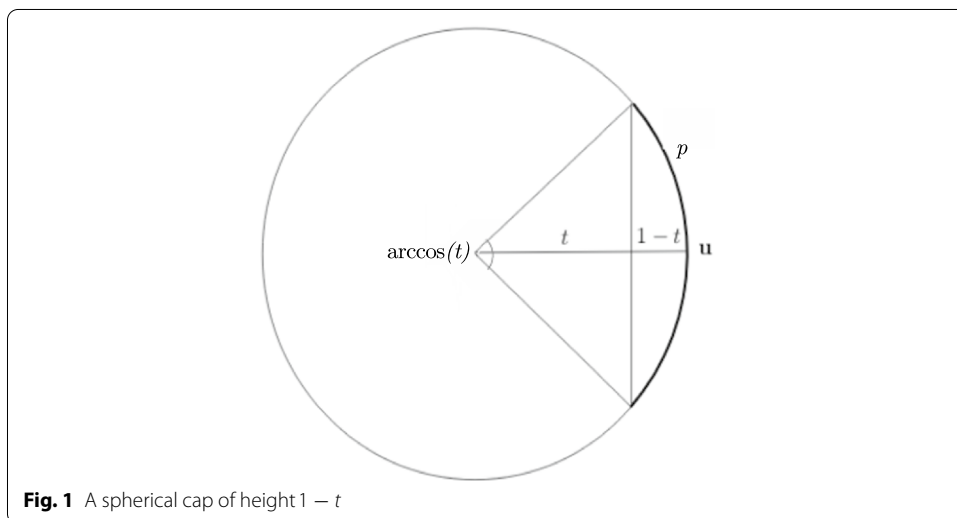


Fig. 1 A spherical cap of height $1 - t$

therefore, it is sufficient to derive a bound on $e^{-t_{p,d}^2(d-1)/2}$. We next show that this quantity might be approximated by $\frac{1}{t_{p,d}\sqrt{d}}(1 - t_{p,d}^2)^{\frac{d-1}{2}}$. Indeed,

$$\frac{1}{t_{p,d}\sqrt{d}}(1 - t_{p,d}^2)^{\frac{d-1}{2}} = \frac{1}{t_{p,d}\sqrt{d}} \exp\left\{\frac{d-1}{2} \log(1 - t_{p,d}^2)\right\},$$

and since $t_{p,d} \rightarrow 0$, we can approximate the logarithm by Taylor series with $k = \lceil 1/\epsilon \rceil + 1$ terms:

$$\log(1 - t_{p,d}^2) = -t_{p,d}^2 - \sum_{i=2}^k \frac{t_{p,d}^{2i}}{i} + O(t_{p,d}^{2k+2}).$$

That implies:

$$\frac{1}{t_{p,d}\sqrt{d}}(1 - t_{p,d}^2)^{\frac{d-1}{2}} = \frac{1}{t_{p,d}\sqrt{d}} \exp\left\{-\frac{d-1}{2} \left(t_{p,d}^2 + \sum_{i=2}^k \frac{t_{p,d}^{2i}}{i} + O(t_{p,d}^{2k+2})\right)\right\}.$$

The first term in the sum under the exponent function gives exactly what we needed. Therefore, it can be expressed as follows:

$$\frac{e^{-t_{p,d}^2(d-1)/2}}{t_{p,d}\sqrt{d}} = \frac{1}{t_{p,d}\sqrt{d}}(1 - t_{p,d}^2)^{\frac{d-1}{2}} \exp\left(\frac{d-1}{2} O(t_{p,d}^{2k+2})\right) \prod_{i=2}^k \exp\left(\frac{(d-1)t_{p,d}^{2i}}{2i}\right). \tag{2}$$

The first term in the right-hand side product, accordingly to (1), is at most proportional to p :

$$\frac{1}{t_{p,d}\sqrt{d}}(1 - t_{p,d}^2)^{\frac{d-1}{2}} \leq 6p.$$

Further, as we learn from Lemma 1, there exists such a constant $C > 0$ that

$$t_{p,d}^{2k+2} \frac{d-1}{2} \leq C \frac{\log^{k+1}(1/p)}{d^k} \leq Cm^{k+1} \frac{\log^{k+1} n}{\log^{k+\epsilon k} n} = Cm^{k+1} \log^{1-\epsilon k} n$$

for large enough n . The right-hand side converges to 0 as $n \rightarrow \infty$ since $1 - \epsilon k < 0$ due to the choice of parameter k . That allows us to bound the second term by a constant larger than 1, for instance:

$$\exp\left(\frac{d-1}{2} O(t_{p,d}^{2k+2})\right) \leq 2.$$

Let us consider the last term of (2). As in the above argumentation,

$$t_{p,d}^{2i} \frac{d-1}{2i} \leq Cm^i \frac{\log^i n}{\log^{i+\epsilon i} n} = Cm^i \log^{1-\epsilon i} n.$$

But $\exp(Cm^k \log^{1-\epsilon i} n)$ grows slower than $n^{\gamma/k}$ for any $\gamma > 0$, which means that for large n the third term of (2) is at most n^γ :

$$\prod_{i=2}^k \exp\left(\frac{(d-1)t_{p,d}^{2i}}{2i}\right) \leq (n^{\gamma/k})^{k-1} \leq \frac{n^{\gamma}}{12}.$$

A final combination of these three bounds finishes the proof:

$$\frac{1}{t_{p,d}\sqrt{d}} \exp\left\{\frac{d-1}{2} \log(1-t_{p,d}^2)\right\} \leq 12p \cdot \frac{n^{\gamma}}{12} \leq pn^{\gamma}.$$

□

Now we can present our result that states that $G(n, p, d)$ does not contain cliques of size 4 almost surely. Let us remind that $N_k(n, p, d)$ denotes the number of k -cliques in $G(n, p, d)$.

Theorem 4 *Suppose that $k \geq 4, n^{-m} \leq p \leq n^{-2/3-\gamma}$ with some $m > 0$ and $\gamma > 0$ and $d \gg \log^{1+\epsilon} n$ with some $\epsilon > 0$. Then*

$$\mathbb{P}[N_k(n, p, d) \geq 1] \rightarrow 0, \quad n \rightarrow \infty.$$

Proof First, let us check that the conditions of Theorem 2 are satisfied with the parameters of our case. It is known that

$$\Phi(x) = 1 + e^{-x^2/2} \left(-\frac{1}{\sqrt{2\pi}x} + O\left(\frac{1}{x^3}\right) \right) \text{ as } x \rightarrow \infty. \tag{3}$$

Thus, in our case

$$\hat{p} = 1 - \Phi(t_{p,d}\sqrt{d}) = \frac{e^{-(t_{p,d}\sqrt{d})^2/2}}{t_{p,d}\sqrt{2\pi d}} (1 + o(1))$$

Therefore, since $t_{p,d}\sqrt{d} \rightarrow \infty$,

$$\begin{aligned} \beta &= 2\sqrt{\log \frac{4}{\hat{p}}} = 2\sqrt{-\log\left(\frac{e^{-t_{p,d}^2 d/2}}{t_{p,d}\sqrt{2\pi d}}(1 + o(1))\right)} \\ &= 2\sqrt{(t_{p,d}\sqrt{d})^2/2 + \log(t_{p,d}\sqrt{2\pi d}) - \log 4 + o(1)} \\ &= \sqrt{2}\sqrt{(t_{p,d}\sqrt{d})^2(1 + o(1))} \leq 2t_{p,d}\sqrt{d} \leq \\ &\leq 2C\sqrt{\log(1/p)} \leq 2C\sqrt{m}\sqrt{\log n}. \end{aligned}$$

Here we have used the result of Lemma 2. Considering $d \gg \log^{1+\epsilon} n$, we obtain that $\beta \rightarrow 0$ as $n \rightarrow \infty$, that is why for sufficiently large n ,

$$\beta\sqrt{\frac{k}{d}} < 1.$$

Now we need to select an appropriate parameter δ_n . Since $\alpha = \sqrt{1 - \beta\sqrt{k/d}} \rightarrow 1$ as $n \rightarrow \infty$, we can take $\delta_n = \log^{1/2-\epsilon/2} n$ for sufficiently large n .

It is only left to check the condition

$$d > \frac{8(k+1)^2 \log \frac{1}{\tilde{p}}}{\delta_n^2} \left(k \log \frac{4}{\tilde{p}} + \log \frac{k-1}{2} \right). \tag{4}$$

Indeed, as far as k is constant, for n large enough

$$\begin{aligned} \frac{8(k+1)^2 \log \frac{1}{\tilde{p}}}{\delta_n^2} \left(k \log \frac{4}{\tilde{p}} + \log \frac{k-1}{2} \right) &\leq \frac{32k(k+1)^2 \log^2 \frac{1}{\tilde{p}}}{\log^{1-\epsilon} n} \\ &\leq \frac{32m^2 k(k+1)^2 \log^2 n}{\log^{1-\epsilon} n} \\ &\leq 32m^2 k(k+1)^2 \log^{1+\epsilon} n. \end{aligned}$$

But d grows faster than $\log^{1+\epsilon} n$, and condition (4) is then satisfied. Thus, it is now possible to apply the bound from Theorem 2.

According to the asymptotic representation of $\Phi(x)$ and Lemma 2, for sufficiently large n ,

$$\begin{aligned} 1 - \Phi(\alpha t_{p,d} \sqrt{d} - \delta_n) &\leq \frac{C_\Phi}{t_{p,d} \sqrt{d}} e^{-(\alpha t_{p,d} \sqrt{d} - \delta_n)^2 / 2} \\ &\leq \frac{C_\Phi}{t_{p,d} \sqrt{d}} e^{-\alpha^2 t_{p,d}^2 d / 2} e^{\alpha t_{p,d} \sqrt{d} \delta_n} e^{-\delta_n^2} \\ &\leq \frac{C_\Phi}{t_{p,d} \sqrt{d}} e^{-(1-\beta \sqrt{k/d}) t_{p,d}^2 d / 2} e^{\alpha t_{p,d} \sqrt{d} \delta_n} \\ &\leq C_\Phi p n^{\gamma/4} e^{\beta \sqrt{kd} t_{p,d}^2 / 2} e^{\alpha t_{p,d} \sqrt{d} \delta_n}, \end{aligned}$$

with some universal constant $C_\Phi > 0$. Since $\beta \leq 2C\sqrt{m}\sqrt{\log n}$ and $t_{p,d} \leq C\sqrt{\frac{\log(1/p)}{d}}$, we have

$$\begin{aligned} \exp\left(\beta \sqrt{kd} t_{p,d}^2 / 2\right) &\leq \exp\left(C\sqrt{km}\sqrt{\log n} t_{p,d}^2 \sqrt{d}\right) \\ &\leq \exp\left(C^3 \sqrt{km} \frac{\log(1/p)}{\sqrt{d}}\right) \leq \\ &\leq \exp\left(\sqrt{k} C^3 m^{3/2} \frac{\log^{3/2} n}{\log^{1/2+\epsilon/2} n}\right) \\ &= \exp\left(\sqrt{k} C^3 m^{3/2} \log^{1-\epsilon/2} n\right). \end{aligned}$$

The second exponent can be bounded similarly:

$$\begin{aligned} \exp\left(\alpha t_{p,d} \sqrt{d} \delta_n\right) &\leq \exp\left(t_{p,d} \sqrt{d} \delta_n\right) \leq \exp\left(C\sqrt{m} \log^{1/2} n \log^{1/2-\epsilon/2} n\right) \\ &= \exp\left(C\sqrt{m} \log^{1-\epsilon/2} n\right). \end{aligned}$$

Therefore, for sufficiently large n ,

$$e^{\beta\sqrt{K/d}t_{p,d}^2d/2}e^{\alpha t_{p,d}\sqrt{d}\delta_n} \leq \exp\left((C\sqrt{m} + \sqrt{k}C^3m^{3/2})\log^{1-\epsilon/2}n\right) < n^{\gamma/4}.$$

Finally, we get that

$$1 - \Phi(\alpha t_{p,d}\sqrt{d} - \delta_n) \leq C_{\Phi}pn^{\gamma/2} \leq C_{\Phi}n^{-2/3-\gamma}n^{\gamma/2} = C_{\Phi}n^{-2/3-\gamma/2}.$$

Let us notice that $\binom{n}{k} \leq \frac{n^k}{k!}$. It is easy to verify that $k - \left(\frac{2}{3} + \gamma\right)\frac{k(k-1)}{2} < 0$ for $k \geq 4$ and $\gamma > 0$. Then for $k \geq 4$,

$$\begin{aligned} \mathbb{E}N_k &\leq e^{1/\sqrt{2}}\binom{n}{k}\left(1 - \Phi(\alpha t_{p,d}\sqrt{d} - \delta_n)\right)\binom{k}{2} \\ &\leq \frac{C_{\Phi}}{k!}e^{1/\sqrt{2}}n^kn^{-(2/3+\gamma)\frac{k(k-1)}{2}} \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

It only remains to mention that

$$\mathbb{P}(N_k \geq 1) \leq \mathbb{E}N_k \rightarrow 0, \quad n \rightarrow \infty.$$

The theorem is proved. □

Number of triangles in the sparse regime: $d \gg \log^3 n$

As noted in the previous section, in the sparse regime, $G(n, p, d)$ does not contain any complete subgraph larger than a triangle. The natural question arises, how many triangles are in $G(n, p, d)$. The next two results give some idea of the expected number of triangles. The first result refers to the case $d \gg \log^3 n$; in this case, the average number of triangles grows as the function $\theta(n)$ that determines the probability $p(n) = \theta(n)/n$.

Our first goal is to obtain a more accurate analogue of Lemma 2.

Lemma 3 *Assume $p \geq n^{-m}$ for some $m > 0$ and $d \gg \log^2 n$. Then the following inequality holds true:*

$$\frac{e^{-t_{p,d}^2d/2}}{p} \leq t_{p,d}\sqrt{d} \leq \frac{12e^{-t_{p,d}^2d/2}}{p}. \tag{5}$$

Proof From (1) we learn that

$$\frac{2(1 - t_{p,d}^2)^{d/2}}{p\sqrt{1 - t_{p,d}^2}} = \frac{2(1 - t_{p,d}^2)^{(d-1)/2}}{p} \leq t_{p,d}\sqrt{d} \leq \frac{6(1 - t_{p,d}^2)^{(d-1)/2}}{p} = \frac{6(1 - t_{p,d}^2)^{d/2}}{p\sqrt{1 - t_{p,d}^2}}.$$

Let us write the Taylor series of $(1 - t_{p,d}^2)^{d/2}$:

$$\begin{aligned} (1 - t_{p,d}^2)^{d/2} &= \exp\left(\frac{d}{2} \ln(1 - t_{p,d}^2)\right) = \exp\left(-\frac{d}{2}(t_{p,d}^2 + O(t_{p,d}^4))\right) \\ &= \exp\left(-\frac{d}{2}t_{p,d}^2\right) \exp\left(O(dt_{p,d}^4)\right). \end{aligned}$$

Lemma 1 and the condition $d \gg \log^2 n$ guarantee that $t_{p,d}^4 d \rightarrow 0$ as $n \rightarrow \infty$. This means that for any $\delta > 0$ and sufficiently large n , the quantity $\exp\left(O(dt_{p,d}^4)\right)$ can be bounded as follows:

$$1 - \delta \leq \exp\left(O(dt_{p,d}^4)\right) \leq 1 + \delta.$$

The same statement holds true for $1/\sqrt{1 - t_{p,d}^2}$:

$$1 - \delta \leq \frac{1}{\sqrt{1 - t_{p,d}^2}} \leq 1 + \delta.$$

Therefore, taking $\delta < 1 - 1/\sqrt{2}$,

$$\frac{e^{-t_{p,d}^2 d/2}}{p} \leq \frac{2(1 - \delta)^2 e^{-t_{p,d}^2 d/2}}{p} \leq t_{p,d} \sqrt{d} \leq \frac{6(1 + \delta)^2 e^{-t_{p,d}^2 d/2}}{p} \leq \frac{12e^{-t_{p,d}^2 d/2}}{p}.$$

Thus, inequality (5) is proved. □

Theorem 5 *Let us suppose that $d \gg \log^3 n$ and $p = \theta(n)/n$ with $n^m \leq \theta(n) \ll n$ for some $m > 0$. Then for any $0 < \epsilon < 1$ and sufficiently large n , the expected number of triangles can be bounded as follows:*

$$\frac{2}{15(2\pi)^{3/2}} (1 - \epsilon)\theta^3(n) \leq \mathbb{E}[N_3(n, p, d)] \leq \frac{288}{(2\pi)^{3/2}} e^{1/\sqrt{2}} (1 + \epsilon)\theta(n)^3.$$

Proof The idea of the proof is quite similar to that of Theorem 4, but it uses both Theorems 1 and 2. Besides, we need more accurate asymptotic analysis, as now a rough bound of Lemma 2 is not sufficient for the application of Theorems 1 and 2. We are going to use more precise Lemma 3.

Upper bound As previously, we first need to verify the conditions of Theorem 2. It is obvious that still

$$\beta \sqrt{\frac{3}{d}} < 1.$$

Take $\delta_n = \frac{\log(1+\epsilon/4)}{C\sqrt{m}\sqrt{\log n}}$. Then Theorem 2 holds true for

$$d \geq \frac{384 \log \frac{1}{p} \log \frac{4}{p}}{\delta_n^2}.$$

But the right-hand side does not grow faster than $\log^3 n$ due to the choice of δ_n and the argumentation similar to that of Theorem 4. Consequently, the above condition is satisfied if only $d \gg \log^3 n$, and now we can apply Theorem 2.

Let us rewrite the bound from this theorem for $k = 3$:

$$\mathbb{E}N_3 \leq e^{1/\sqrt{2}} \binom{n}{3} \left(1 - \Phi(\alpha t_{p,d} \sqrt{d} - \delta_n)\right)^3.$$

Of course, the most important term here is $1 - \Phi(\alpha t_{p,d} \sqrt{d} - \delta_n)$. Similar to the proof of the previous theorem one can get (with the asymptotic representation (3)) that

$$1 - \Phi(\alpha t_{p,d} \sqrt{d} - \delta_n) \leq \frac{e^{-t_{p,d}^2 d/2}}{t_{p,d} \sqrt{2\pi d}} e^{\beta \sqrt{3d} t_{p,d}^2/2} e^{\alpha t_{p,d} \sqrt{d} \delta_n}.$$

From (5) we learn that

$$\frac{e^{-t_{p,d}^2 d/2}}{t_{p,d} \sqrt{d}} \leq 12p.$$

Further, since $d \gg \log^3 n$ and $\beta < 2C\sqrt{m}\sqrt{\log n}$ (see the proof of Theorem 4), for sufficiently large n ,

$$\begin{aligned} \exp\left(\beta \sqrt{3d} t_{p,d}^2\right) &\leq \exp\left(\frac{2\sqrt{3}C^3 \sqrt{m} \sqrt{\log n} \log(1/p)}{\sqrt{d}}\right) \\ &\leq \exp\left(\frac{2\sqrt{3}C^3 m^{3/2} \log^{3/2} n}{\sqrt{d}}\right) = 1 + o(1), \quad n \rightarrow \infty. \end{aligned}$$

The second exponent is just a constant for the chosen δ_n (here we use the fact that $\alpha < 1$):

$$\exp\left(\alpha t_{p,d} \sqrt{d} \delta_n\right) \leq \exp\left(t_{p,d} \sqrt{d} \delta_n\right) \leq \exp\left(\sqrt{Cm} \delta_n \log^{1/2} n\right) = e^{\log(1+\epsilon/2)} = 1 + \epsilon/2.$$

Putting all together and taking into account that $\binom{n}{3} \leq \frac{n^3}{6}$,

$$\begin{aligned} \mathbb{E}[N_3(n, p, d)] &\leq e^{1/\sqrt{2}} \binom{n}{3} \left(1 - \Phi(\alpha t_{p,d} \sqrt{d} - \delta_n)\right)^3 \\ &\leq \frac{288}{(2\pi)^{3/2}} e^{1/\sqrt{2}} (1 + \epsilon/2)(1 + o(1)) n^3 p^3 \\ &= \frac{288}{(2\pi)^{3/2}} e^{1/\sqrt{2}} (1 + \epsilon) \theta(n)^3. \end{aligned}$$

Lower bound Now we are going to use Theorem 1. First, we need to determine the asymptotic behavior of the function \tilde{p} :

$$\tilde{p} = 1 - \Phi(2t_{p,d} \sqrt{d} + 1) = \frac{e^{-2t_{p,d}^2 \sqrt{d}}}{2t_{p,d} \sqrt{2\pi d}} (1 + o(1)).$$

As one can easily see,

$$\log \frac{1}{\tilde{p}} = \log\left(2\sqrt{2\pi d} t_{p,d} e^{t_{p,d}^2 d} (1 + o(1))\right) = t_{p,d}^2 d + \log\left(2\sqrt{2\pi d} t_{p,d}\right) = 2 \log n (1 + o(1)).$$

Then $\beta = \Theta(\sqrt{\log n})$ and $\beta\sqrt{3/d} \rightarrow 0$ with $n \rightarrow \infty$. This implies that $\alpha = \sqrt{1 - \beta\sqrt{k/d}} \rightarrow 1$ as $n \rightarrow \infty$. Let us take $\delta_n = \frac{\log(1+\epsilon)}{2\alpha C\sqrt{m}\sqrt{\log n}}$. Similar to the previous case, the condition

$$d \geq \frac{384 \log \frac{1}{p} \log \frac{4}{p}}{\delta_n^2}$$

is satisfied if $d \gg \log^3 n$.

Let us remind that the bound of Theorem 1 is written as follows:

$$\mathbb{E}[N_3(n, p, d)] \geq \frac{4}{5} \binom{n}{3} (1 - \tilde{\Phi}_3(d, p)) \binom{3}{2},$$

where $\tilde{\Phi}_3(d, p) = \Phi\left(\frac{\alpha t_{p,d}\sqrt{d} + \delta_n}{\sqrt{1 - \frac{32 \log(1/\tilde{p})}{d}}}\right)$.

Since $t_{p,d}\sqrt{d} \rightarrow \infty$ and $\alpha \rightarrow 1$ as $n \rightarrow \infty$,

$$\begin{aligned} 1 - \tilde{\Phi}_3(d, p) &= \frac{1}{t_{p,d}\sqrt{2\pi d}} \exp\left\{-\frac{(\alpha t_{p,d}\sqrt{d} + \delta_n)^2}{1 - \frac{32 \log(1/\tilde{p})}{d}}\right\} (1 + o(1)) \\ &\geq \frac{1}{t_{p,d}\sqrt{2\pi d}} \exp\left\{-\frac{(\alpha t_{p,d}\sqrt{d} + \delta_n)^2}{d} \left(1 + \frac{64 \log(1/\tilde{p})}{d}\right)\right\} (1 + o(1)). \end{aligned}$$

Here we used a simple inequality $1/(1-x) < 1 + 2x$ for $0 < x < 1/2$ and the fact that $\frac{4(K+1)^2 \log(1/\tilde{p})}{d} = O(1/\log^2 n) \rightarrow 0$ as $n \rightarrow \infty$. By the same reason,

$$(\alpha t_{p,d}\sqrt{d} + \delta_n)^2 \frac{4(K+1)^2 \log(1/\tilde{p})}{d} = O\left(\frac{1}{\log n}\right).$$

Hence,

$$\exp\left\{-\frac{(\alpha t_{p,d}\sqrt{d} + \delta_n)^2 \frac{4(K+1)^2 \log(1/\tilde{p})}{d}}{d}\right\} = 1 + o(1), \quad n \rightarrow \infty.$$

Consequently,

$$\begin{aligned} 1 - \tilde{\Phi}_3(d, p) &= \frac{1}{t_{p,d}\sqrt{2\pi d}} \exp\left\{-\frac{(\alpha t_{p,d}\sqrt{d} + \delta_n)^2}{d}\right\} (1 + o(1)) \\ &= \frac{1}{t_{p,d}\sqrt{2\pi d}} \exp\left\{-\frac{(\alpha^2 t_{p,d}^2 d + 2\alpha t_{p,d}\delta_n\sqrt{d} + \delta_n^2)}{d}\right\} (1 + o(1)) \\ &= \frac{1}{t_{p,d}\sqrt{2\pi d}} \exp\left\{-\left(1 + \sqrt{\frac{8K}{d} \log \frac{4}{p}}\right) t_{p,d}^2 d - 2\alpha t_{p,d}\delta_n\sqrt{d} - \delta_n^2\right\} (1 + o(1)). \end{aligned}$$

The inequality (5) guarantees that

$$\frac{e^{-t_{p,d}^2 d}}{t_{p,d}\sqrt{d}} \geq p.$$

Further, similarly to the previous case,

$$\exp(-2\alpha t_{p,d}\sqrt{d}\delta_n) \geq \exp(-\log(1 + \epsilon/4)) = 1/(1 + \epsilon/4) \geq 1 - \epsilon/2.$$

Finally, it is easy to check that $\delta_n^2 \rightarrow 0$ and $\sqrt{\frac{8K}{d} \log \frac{4}{p} t_{p,d}^2} d \rightarrow 0$ under the condition $d \gg \log^3 n$. That is why

$$1 - \tilde{\Phi}_3(d, p) \geq \frac{p}{\sqrt{2\pi}}(1 - \epsilon/2)(1 + o(1)) \geq \frac{p}{\sqrt{2\pi}}(1 - \epsilon).$$

That leads us to the final bound:

$$\begin{aligned} \mathbb{E}N_3(n, p, d) &\geq \frac{4}{5} \binom{n}{3} \left(1 - \tilde{\Phi}_3(d, p)\right)^3 \geq \frac{2}{15(2\pi)^{3/2}}(1 - \epsilon)n^3 p^3 \\ &= \frac{2}{15(2\pi)^{3/2}}(1 - \epsilon)\theta^3(n). \end{aligned}$$

□

Number of triangles in the sparse regime: $d \ll \log^3 n$

So far, the presented results more likely confirm the similarity of random geometric graphs and Erdős–Rényi graphs. However, from Bubeck et al. (2016) one can learn that these graphs are completely different in the sparse regime if $d \ll \log^3 n$. This can be easily deduced from the result of Theorem 3. It states that the expected number of triangles of a random geometric graph grows significantly faster (as a polylogarithmic function of n) than one of the corresponding Erdős–Rényi graph. It turns out that the bound of Theorem 3 can be improved.

In order to make this improvement, we present some results from convex geometry. First of all, it is known that the surface area A_d of $(d - 1)$ -dimensional sphere \mathbb{S}^{d-1} can be calculated as follows (see Blumenson 1960):

$$A_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}.$$

Now we need a result providing the expression for the surface area of the intersection of two spherical caps in \mathbb{R}^d . Let us denote by $A_d(\theta_1, \theta_2, \theta_v)$ the surface area of the intersection of two spherical caps of angles θ_1 and θ_2 with the angle θ_v between axes defining these caps. The paper (Lee and Kim 2014) gives the exact formula for this quantity in terms of the regularized incomplete beta function.

Theorem 6 (Lee and Kim 2011) *Let us suppose that $\theta_v \in [0, \pi/2)$ and $\theta_1, \theta_2 \in [0, \theta_v]$. Then*

$$\begin{aligned}
 A_d(\theta_1, \theta_2, \theta_v) &= \frac{\pi^{(n-1)/2}}{\Gamma\left(\frac{n-1}{2}\right)} \left\{ \int_{\theta_{min}}^{\theta_2} \sin^{d-2} \phi I_{1-\left(\frac{\tan \theta_{min}}{\tan \phi}\right)^2} \left(\frac{n-1}{2}, \frac{1}{2}\right) d\phi \right. \\
 &\quad \left. + \int_{\theta_v-\theta_{min}}^{\theta_1} \sin^{d-2} \phi I_{1-\left(\frac{\tan(\theta_v-\theta_{min})}{\tan \phi}\right)^2} \left(\frac{n-1}{2}, \frac{1}{2}\right) d\phi \right\} \\
 &:= J_n^{\theta_{min}, \theta_2} + J_n^{\theta_v-\theta_{min}, \theta_1},
 \end{aligned}$$

where θ_{min} is defined as follows

$$\theta_{min} = \arctan \left(\frac{\cos \theta_1}{\cos \theta_2 \sin \theta_v} - \frac{1}{\tan \theta_v} \right),$$

and $I_x(a, b)$ stands for the regularized incomplete beta function, that is

$$I_x(a, b) = \frac{B(x, a, b)}{B(a, b)} = \frac{\int_0^x t^{a-1}(1-t)^{b-1} dt}{\int_0^1 t^{a-1}(1-t)^{b-1} dt}.$$

Theorem 7 Let $d \gg \log^2 n$, and assume $p = \theta(n)/n$ with $n^m \leq \theta(n) \ll n$ for some $m > 0$. Then there exist constants $C_l > 0$ and $C_u > 0$ such that

$$C_l \theta^3(n) t_{p,d}^2 e^{t_{p,d}^3} (1 + o(1)) \leq \mathbb{E}[N_3(n, p, d)] \leq C_u \theta^3(n) e^{t_{p,d}^3} (1 + o(1)).$$

Proof **Notation and a general plan of the proof** Let us make some preparations. Denote by $E_{i,j}$ the event $\{\langle X_i, X_j \rangle \geq t_{p,d}\}$ and by $E_{i,j}(x)$ the event $\{\langle X_i, X_j \rangle = x\}$. In what follows, we condition on the zero probability event $E_{i,j}(x)$. It should be understood as conditioning on the event $\{x - \epsilon \leq \langle X_i, X_j \rangle \leq x + \epsilon\}$ with $\epsilon \rightarrow 0$. Using this notation, we can rewrite

$$\begin{aligned}
 \mathbb{P}(E_{1,2}E_{1,3}E_{2,3}) &= \int_{t_{p,d}}^1 \mathbb{P}(E_{2,3}E_{1,3}|E_{1,2}(x)) f_d(x) \\
 &= \int_{t_{p,d}}^{2t_{p,d}} \mathbb{P}(E_{2,3}E_{1,3}|E_{1,2}(x)) f_d(x) dx + \int_{2t_{p,d}}^1 \mathbb{P}(E_{2,3}E_{1,3}|E_{1,2}(x)) f_d(x) dx \\
 &:= T_1 + T_2,
 \end{aligned} \tag{6}$$

where $f_d(x)$ is the density of a coordinate of a uniform random point on \mathbb{S}^{d-1} (see Bubeck et al. 2016), that is

$$f_d(x) = \frac{\Gamma(d/2)}{\Gamma((d-1)/2)\sqrt{\pi}} (1-x^2)^{(d-3)/2}, \quad x \in [-1, 1].$$

Using the fact that

$$\Gamma(d)\sqrt{d}/100 \leq \Gamma(d+1/2) \leq 2\sqrt{d}\Gamma(d), \tag{7}$$

we can present $f_d(x)$ as follows:

$$f_d(x) = C_f(d)\sqrt{d}(1-x^2)^{(d-3)/2}, \quad x \in [-1, 1]. \tag{8}$$

Here $C_f(d)$ denotes some function of d with $1/100 \leq C_f(d) \leq \sqrt{2}$.

Here is a general outline of the proof. We treat the terms T_1 and T_2 separately and start with T_1 . The probability $P(E_{2,3}E_{1,3}|E_{1,2}(x))$ can be expressed with the normalized surface area of the intersection of two spherical caps. First, we need to bound this quantity. After that, using the representation (8), we will calculate T_1 in terms of the CDF of the standard normal distribution and will estimate its asymptotic behavior. As for T_2 , it will be enough to show that $T_2 = o(T_1)$ as $n \rightarrow \infty$.

Estimation of term T_1 . As was mentioned above, we start with T_1 . It will be more hand-ful to write it in the following form:

$$\begin{aligned} T_1 &= \int_{t_{p,d}}^{2t_{p,d}} \mathbb{P}(E_{2,3}E_{1,3}|E_{1,2}(x))f_d(x)dx = \\ &= t_{p,d} \int_1^2 \mathbb{P}(E_{2,3}E_{1,3}|E_{1,2}(\alpha t_{p,d}))f_d(\alpha t_{p,d})d\alpha. \end{aligned}$$

Conditioning on $E_{1,2}(\alpha t_{p,d})$, the probability $\mathbb{P}(E_{2,3}E_{1,3}|E_{1,2}(\alpha t_{p,d}))$ is just the normalized surface area of the intersection of two caps of angle $\arccos(t_{p,d})$ (and $\arccos(\alpha t_{p,d})$ is the angle between the axes of these caps).

$$\begin{aligned} \tilde{p}(\alpha) &:= P(E_{2,3}E_{1,3}|E_{1,2}(\alpha t_{p,d})) = \\ &= \frac{A_d(\arccos(t_{p,d}), \arccos(t_{p,d}), \arccos(\alpha t_{p,d}))}{A_d} = \\ &= \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^{d/2}} \left(J_d^{\theta_{min}, \arccos(t_{p,d})} + J_d^{\arccos(\alpha t_{p,d}) - \theta_{min}, \arccos(t_{p,d})} \right), \end{aligned} \tag{9}$$

where $J_d^{a,b}$ is defined in Theorem 6, and

$$\theta_{min} = \arctan \left(\frac{1}{\sin(\arccos(\alpha t_{p,d}))} - \frac{1}{\tan(\arccos(\alpha t_{p,d}))} \right) = \arctan \left(\sqrt{\frac{1 - \alpha t_{p,d}}{1 + \alpha t_{p,d}}} \right).$$

Because both caps are of the same angle, the parts in the right-hand side of (9) are equal. Therefore, recalling the definition of $J_d^{\theta_{min}, \arccos(t_{p,d})}$,

$$\begin{aligned} \tilde{p}(\alpha) &= \frac{\Gamma\left(\frac{d}{2}\right)}{\pi^{d/2}} J_d^{\theta_{min}, \arccos(t_{p,d})} \\ &= \frac{\Gamma\left(\frac{d}{2}\right)\pi^{(d-1)/2}}{\pi^{d/2}\Gamma\left(\frac{d-1}{2}\right)} \int_{\theta_{min}}^{\arccos(t_{p,d})} \sin^{d-2} \phi I_{1 - \left(\frac{\tan \theta_{min}}{\tan \phi}\right)^2} \left(\frac{d-2}{2}, \frac{1}{2} \right) d\phi. \end{aligned}$$

Let us make a change of variables: $\sin^2 \phi = z$. Using this change and the expression for θ_{min} , we obtain

$$\tilde{p}(\alpha) = \frac{\Gamma\left(\frac{d}{2}\right)\pi^{(d-1)/2}}{2\pi^{d/2}\Gamma\left(\frac{d-1}{2}\right)} \int_{\frac{1-\alpha t_{p,d}}{2}}^{1-t_{p,d}^2} z^{(d-3)/2} (1-z)^{-1/2} I_{1 - \frac{1-\alpha t_{p,d}}{1+\alpha t_{p,d}} \cdot \frac{1-z}{z}} \left(\frac{d-2}{2}, \frac{1}{2} \right) dz.$$

Considering the definition of a regularized beta function and the formula $\Gamma(d+1) = d\Gamma(d)$, we have

$$\tilde{p}(\alpha) = \frac{d}{2\pi} \int_{\frac{1-\alpha t_{p,d}}{2}}^{1-t_{p,d}^2} z^{(d-3)/2} (1-z)^{-1/2} \left(\int_0^{1-\frac{1-\alpha t_{p,d}}{1+\alpha t_{p,d}} \cdot \frac{1-z}{z}} y^{d/2-2} (1-y)^{-1/2} dy \right) dz.$$

Next, we need a simple double bound on the incomplete beta function $I_u(a, 1/2)$:

$$\frac{u^{a+1}}{a+1} \leq \int_0^u t^a (1-t)^{-1/2} dt \leq \frac{u^{a+1}}{(a+1)\sqrt{1-u}},$$

which can be established by estimation of $(1-t)^{-1/2}$ and subsequent explicit integration. That is why

$$\begin{aligned} & \frac{1}{\pi} \int_{\frac{1-\alpha t_{p,d}}{2}}^{1-t_{p,d}^2} z^{(d-3)/2} (1-z)^{-1/2} \left(1 - \frac{1-\alpha t_{p,d}}{1+\alpha t_{p,d}} \cdot \frac{1-z}{z} \right)^{d/2-1} dz \leq \tilde{p}(\alpha) \\ & \leq \frac{2}{\pi} \int_{\frac{1-\alpha t_{p,d}}{2}}^{1-t_{p,d}^2} z^{(d-3)/2} (1-z)^{-1/2} \left(\frac{1-\alpha t_{p,d}}{1+\alpha t_{p,d}} \cdot \frac{1-z}{z} \right)^{-1/2} \left(1 - \frac{1-\alpha t_{p,d}}{1+\alpha t_{p,d}} \cdot \frac{1-z}{z} \right)^{d/2-1} dz. \end{aligned}$$

Here we used the fact that $1 \leq d/(d-2) \leq 2$ for $d \geq 4$. We can transform:

$$1 - \frac{1-\alpha t_{p,d}}{1+\alpha t_{p,d}} \cdot \frac{1-z}{z} = \frac{2z-1+\alpha t_{p,d}}{(1+\alpha t_{p,d})z},$$

which gives us the following estimation:

$$\begin{aligned} & \frac{1}{2\pi} \int_{\frac{1-\alpha t_{p,d}}{2}}^{1-t_{p,d}^2} \frac{1}{\sqrt{z(1-z)}} \left(\frac{2z-1+\alpha t_{p,d}}{1+\alpha t_{p,d}} \right)^{d/2-1} dz \leq \tilde{p}(\alpha) \\ & \leq \frac{1}{\pi} \sqrt{\frac{1+\alpha t_{p,d}}{1-\alpha t_{p,d}}} \int_{\frac{1-\alpha t_{p,d}}{2}}^{1-t_{p,d}^2} \frac{1}{1-z} \left(\frac{2z-1+\alpha t_{p,d}}{1+\alpha t_{p,d}} \right)^{d/2-1} dz \end{aligned}$$

It is easy to check that $\frac{1}{\sqrt{z(1-z)}}$ has a minimum value $1/2$ for $z \in [0, 1]$, and $\frac{1}{1-z}$ is increasing in z for $z > 0$. Therefore,

$$\begin{aligned} & \frac{1}{2\pi} \int_{\frac{1-\alpha t_{p,d}}{2}}^{1-t_{p,d}^2} \left(\frac{2z-1+\alpha t_{p,d}}{1+\alpha t_{p,d}} \right)^{d/2-1} dz \leq \tilde{p}(\alpha) \\ & \leq \frac{2}{\pi t_{p,d}^2} \sqrt{\frac{1+\alpha t_{p,d}}{1-\alpha t_{p,d}}} \int_{\frac{1-\alpha t_{p,d}}{2}}^{1-t_{p,d}^2} \left(\frac{2z-1+\alpha t_{p,d}}{1+\alpha t_{p,d}} \right)^{d/2-1} dz. \end{aligned}$$

Now we can explicitly compute the integral:

$$\int_{\frac{1-\alpha t_{p,d}}{2}}^{1-t_{p,d}^2} \left(\frac{2z-1+\alpha t_{p,d}}{1+\alpha t_{p,d}} \right)^{d/2-1} dz = \frac{1+\alpha t_{p,d}}{d} \left(1 - \frac{2t_{p,d}^2}{1+\alpha t_{p,d}} \right)^{d/2},$$

which implies the final bounds on $\tilde{p}(\alpha)$:

$$\frac{1+\alpha t_{p,d}}{2\pi d} \left(1 - \frac{2t_{p,d}^2}{1+\alpha t_{p,d}} \right)^{d/2} \leq \tilde{p}(\alpha) \leq \frac{2(1+\alpha t_{p,d})^{3/2}}{\pi d t_{p,d}^2 \sqrt{1-\alpha t_{p,d}}} \left(1 - \frac{2t_{p,d}^2}{1+\alpha t_{p,d}} \right)^{d/2}.$$

This means that T_1 can be estimated as follows:

$$\begin{aligned} & \frac{t_{p,d}}{2\pi d} \int_1^2 (1 + \alpha t_{p,d}) \left(1 - \frac{2t_{p,d}^2}{1 + \alpha t_{p,d}}\right)^{d/2} f_d(\alpha t_{p,d}) d\alpha \leq T_1 \\ & \leq \frac{2}{\pi d t_{p,d}} \int_1^2 \frac{(1 + \alpha t_{p,d})^{3/2}}{\sqrt{1 - \alpha t_{p,d}}} \left(1 - \frac{2t_{p,d}^2}{1 + \alpha t_{p,d}}\right)^{d/2} f_d(\alpha t_{p,d}) d\alpha. \end{aligned} \tag{10}$$

Let us recall (8) and rewrite the ‘essential’ part of previous inequalities.

$$\begin{aligned} & \left(1 - \frac{2t_{p,d}^2}{1 + \alpha t_{p,d}}\right)^{d/2} f_d(\alpha t_{p,d}) \\ & = C_f(d) \sqrt{d} \left(1 - \frac{2t_{p,d}^2}{1 + \alpha t_{p,d}}\right)^{3/2} \left((1 - \alpha^2 t_{p,d}^2) \left(1 - \frac{2t_{p,d}^2}{1 + \alpha t_{p,d}}\right) \right)^{(d-3)/2} \\ & = C_f(d) \sqrt{d} \left(1 - \frac{2t_{p,d}^2}{1 + \alpha t_{p,d}}\right)^{3/2} \left(1 - (2 + \alpha^2)t_{p,d}^2 + 2\alpha t_{p,d}^3\right)^{(d-3)/2} \\ & = C_f(d) \sqrt{d} \left(1 - \frac{2t_{p,d}^2}{1 + \alpha t_{p,d}}\right)^{3/2} \frac{\left(1 - (2 + \alpha^2)t_{p,d}^2 + 2\alpha t_{p,d}^3\right)^{(d-1)/2}}{1 - (2 + \alpha^2)t_{p,d}^2 + 2\alpha t_{p,d}^3}. \end{aligned} \tag{11}$$

Of course, we are most interested in the term $\left(1 - (2 + \alpha^2)t_{p,d}^2 + 2\alpha t_{p,d}^3\right)^{(d-1)/2}$, which can be rewritten in the following form:

$$\left(1 - (2 + \alpha^2)t_{p,d}^2 + 2\alpha t_{p,d}^3\right)^{(d-1)/2} = \exp \left\{ \frac{d-1}{2} \log \left(1 - (2 + \alpha^2)t_{p,d}^2 + 2\alpha t_{p,d}^3\right) \right\}.$$

Since $t_{p,d} \rightarrow 0$ as $n \rightarrow \infty$, one can use Taylor series for logarithm:

$$\begin{aligned} \log \left(1 - (2 + \alpha^2)t_{p,d}^2 + 2\alpha t_{p,d}^3\right) & = -(2 + \alpha^2)t_{p,d}^2 + 2\alpha t_{p,d}^3 + O(t_{p,d}^4) \\ & = -(3 + (\alpha^2 - 1))t_{p,d}^2 + 2\alpha t_{p,d}^3 + O(t_{p,d}^4), \quad n \rightarrow \infty. \end{aligned}$$

From (5) one can easily deduce that

$$\exp \left(-\frac{d-1}{2} t_{p,d}^2 \right) = C_e(p, d) p t_{p,d} \sqrt{d} \exp \left(O(t_{p,d}^4) \right),$$

where $1 \leq C_e(p, d) \leq 12$ is some function that depends only on p and d . Therefore,

$$\begin{aligned} & \left(1 - (3 + (\alpha^2 - 1))t_{p,d}^2 + 2\alpha t_{p,d}^3\right)^{(d-1)/2} \\ & = \exp \left\{ -3 \frac{d-1}{2} t_{p,d}^2 \right\} \exp \left\{ -\frac{d-1}{2} \left((\alpha^2 - 1)t_{p,d}^2 - 2\alpha t_{p,d}^3 + O(t_{p,d}^4) \right) \right\} \\ & = C_e^3(p, d) p^3 t_{p,d}^3 d^{3/2} \exp \left\{ -\frac{d-1}{2} \left((\alpha^2 - 1)t_{p,d}^2 - 2\alpha t_{p,d}^3 + O(t_{p,d}^4) \right) \right\}. \end{aligned} \tag{12}$$

We have transformed the main term of (11). Let us deal with ‘unimportant’ parts of (10) and (11). Denote

$$\begin{aligned}
 h_l(\alpha, p, d) &= \frac{C_f(d)C_e^3(p, d)}{2\pi} \cdot \frac{(1 + \alpha t_{p,d}) \left(1 - \frac{2t_{p,d}^2}{1 + \alpha t_{p,d}}\right)^{3/2}}{1 - (3 + (\alpha^2 - 1))t_{p,d}^2 + 2\alpha t_{p,d}^3}; \\
 h_u(\alpha, p, d) &= \frac{2C_f(d)C_e^3(p, d)}{\pi} \cdot \frac{(1 + \alpha t_{p,d})^{3/2}}{\sqrt{1 - \alpha t_{p,d}}} \cdot \frac{\left(1 - \frac{2t_{p,d}^2}{1 + \alpha t_{p,d}}\right)^{3/2}}{1 - (3 + (\alpha^2 - 1))t_{p,d}^2 + 2\alpha t_{p,d}^3}.
 \end{aligned}$$

Since $1 \leq \alpha \leq 2$ and $t_{p,d} \rightarrow 0$, for sufficiently large n and some constants $C_l > 0$ and $C_u > 0$,

$$h_l(\alpha, p, d) \geq \frac{C_f(d)C_e^3(p, d)}{4\pi} \cdot \frac{(1 + t_{p,d}) \left(1 - \frac{2t_{p,d}^2}{1 + t_{p,d}}\right)^{5/2}}{1 - 3t_{p,d}^2 + 4t_{p,d}^3} \geq 6\sqrt{2\pi} C_l. \tag{13}$$

and

$$h_u(\alpha, p, d) \leq \frac{C_f(d)C_e^3(p, d)(1 + 2t_{p,d})^{3/2} \left(1 - \frac{2t_{p,d}^2}{1 + 2t_{p,d}}\right)^{3/2}}{\pi \sqrt{1 - t_{p,d}}(1 - 6t_{p,d}^2 + 2t_{p,d}^3)} \leq 6\sqrt{2\pi} C_u. \tag{14}$$

Then, plugging (11), (12), (13) and (14) into (10), we obtain the following final bounds on T_1 at this step:

$$\begin{aligned}
 6\sqrt{2\pi} C_l dt_{p,d}^4 p^3 \int_1^2 \exp \left\{ -\frac{d-1}{2} \left((\alpha^2 - 1)t_{p,d}^2 - 2\alpha t_{p,d}^3 + O(t_{p,d}^4) \right) \right\} d\alpha &\leq T_1 \leq \\
 \leq 6\sqrt{2\pi} C_u dt_{p,d}^2 p^3 \int_1^2 \exp \left\{ -\frac{d-1}{2} \left((\alpha^2 - 1)t_{p,d}^2 - 2\alpha t_{p,d}^3 + O(t_{p,d}^4) \right) \right\} d\alpha. & \tag{15}
 \end{aligned}$$

Expression of bounds on T_1 with the CDF of the standard normal distribution One can easily get that

$$(\alpha^2 - 1)t_{p,d}^2 - 2\alpha t_{p,d}^3 + O(t_{p,d}^4) = t_{p,d}^2 \left((\alpha - t_{p,d})^2 - 1 + O(t_{p,d}^2) \right),$$

which implies

$$\begin{aligned}
 &\int_1^2 \exp \left\{ -\frac{d-1}{2} \left((\alpha^2 - 1)t_{p,d}^2 - 2\alpha t_{p,d}^3 + O(t_{p,d}^4) \right) \right\} d\alpha \\
 &= \exp \left\{ \frac{d-1}{2} t_{p,d}^2 (1 + O(t_{p,d}^2)) \right\} \int_1^2 \exp \left\{ -\frac{d-1}{2} t_{p,d}^2 (\alpha - t_{p,d})^2 \right\} d\alpha.
 \end{aligned}$$

Let us treat the integral in the right-hand side of the last equation. Changing the variable $\beta = (\alpha - t_{p,d})t_{p,d}\sqrt{d-1}$, we obtain that

$$\int_1^2 \exp \left\{ -\frac{d-1}{2} t_{p,d}^2 (\alpha - t_{p,d})^2 \right\} d\alpha = \frac{1}{t_{p,d} \sqrt{d-1}} \int_{(1-t_{p,d})t_{p,d}\sqrt{d-1}}^{(2-t_{p,d})t_{p,d}\sqrt{d-1}} e^{-\beta^2/2} d\beta$$

$$= \frac{1}{t_{p,d} \sqrt{d-1}} \left(\Phi((2-t_{p,d})t_{p,d}\sqrt{d-1}) - \Phi((1-t_{p,d})t_{p,d}\sqrt{d-1}) \right).$$

Since $t_{p,d}\sqrt{d-1} \rightarrow \infty$ as $n \rightarrow \infty$,

$$\frac{\Phi((2-t_{p,d})t_{p,d}\sqrt{d-1}) - \Phi((1-t_{p,d})t_{p,d}\sqrt{d-1})}{(1-t_{p,d})t_{p,d}\sqrt{2\pi(d-1)}} (1 + o(1)) - \frac{e^{-(2-t_{p,d})^2 t_{p,d}^2 (d-1)/2}}{(2-t_{p,d})t_{p,d}\sqrt{2\pi(d-1)}} (1 + o(1)).$$

But the ratio of the second and the first terms in the right-hand side converges to 0 as $n \rightarrow \infty$. Indeed,

$$\frac{e^{-(2-t_{p,d})^2 t_{p,d}^2 (d-1)/2}}{e^{-(1-t_{p,d})^2 t_{p,d}^2 (d-1)/2}} \cdot \frac{2-t_{p,d}}{1-t_{p,d}} \leq 3 \exp \left\{ \frac{d-1}{2} t_{p,d}^2 ((1-t_{p,d})^2 - (2-t_{p,d})^2) \right\} \tag{16}$$

$$= 3 \exp \left\{ \frac{d-1}{2} t_{p,d}^2 (-3 + 2t_{p,d}) \right\} \leq 3 \exp \left\{ -2 \cdot \frac{d-1}{2} t_{p,d}^2 \right\} \rightarrow 0, \quad n \rightarrow \infty.$$

The condition $t_{p,d} \rightarrow 0$ implies that

$$\int_1^2 \exp \left\{ -\frac{d-1}{2} t_{p,d} (\alpha - t_{p,d})^2 \right\} d\alpha = \frac{e^{-(1-t_{p,d})^2 t_{p,d}^2 (d-1)/2}}{t_{p,d}^2 d \sqrt{2\pi}} (1 + o(1)),$$

and, therefore,

$$\exp \left\{ \frac{d-1}{2} t_{p,d}^2 (1 + O(t_{p,d}^2)) \right\} \int_1^2 \exp \left\{ -\frac{d-1}{2} t_{p,d} (\alpha - t_{p,d})^2 \right\} d\alpha$$

$$= \frac{e^{d(t_{p,d}^3 + O(t_{p,d}^4))}}{t_{p,d}^2 d \sqrt{2\pi}} (1 + o(1)) = \frac{e^{dt_{p,d}^3}}{t_{p,d}^2 d \sqrt{2\pi}} (1 + o(1)). \tag{17}$$

The last equality holds because under the condition $d \gg \log^2 n$, it is true that $dt_{p,d}^4 \rightarrow 0$ as $n \rightarrow \infty$, and $e^{dt_{p,d}^4} \rightarrow 1$. Putting (17) in (15) gives the final bounds on T_1 :

$$6C_l p^3 t_{p,d}^2 e^{t_{p,d}^3 d} (1 + o(1)) \leq T_1 \leq 6C_u p^3 e^{t_{p,d}^3 d} (1 + o(1)).$$

This concludes the first part the proof.

Estimation of T_2 The second term can be treated much more easily. Indeed, let us bound from above:

$$\begin{aligned} T_2 &= \int_{2t_{p,d}}^1 \mathbb{P}(E_{2,3}E_{1,3}|E_{1,2}(x))f_d(x)dx \leq \int_{2t_{p,d}}^1 f_d(x)dx \\ &= C_f(d)\sqrt{d} \int_{2t_{p,d}}^1 (1-x^2)^{(d-3)/2}dx \leq C_f(d)\sqrt{d}(1-4t_{p,d}^2)^{(d-3)/2} \\ &\leq C_f(d)\sqrt{d}e^{-4t_{p,d}^2(d-3)/2}. \end{aligned}$$

But, similarly to the argumentation in (16), $\sqrt{d}e^{-4t_{p,d}^2(d-3)/2}$ is $o(p^3t_{p,d}^2e^{t_{p,d}^3d})$, hence, finally,

$$6C_l p^3 t_{p,d}^2 e^{t_{p,d}^3 d} (1 + o(1)) \leq T_1 + T_2 = \mathbb{P}(E_{1,2}E_{2,3}E_{1,3}) \leq 6C_u p^3 e^{t_{p,d}^3 d} (1 + o(1)).$$

It is only left to use the standard asymptotic representation of the binomial coefficient $\binom{n}{3} = \frac{n^3}{6} (1 + o(1))$ in order to obtain the bounds on the expected number of triangles:

$$C_l \theta^3(n) t_{p,d}^2 e^{t_{p,d}^3 d} (1 + o(1)) \leq \mathbb{E}[N_3(n, p, d)] \leq C_u \theta^3(n) e^{t_{p,d}^3 d} (1 + o(1)), \quad n \rightarrow \infty.$$

The theorem is proved. □

Let us now discuss the result of this theorem. To make the expressions more handfull, we consider only $p = 1/n$, but the idea can be extended up to any sufficiently small p . First of all, in this case, as we know from Lemma 1, $t_{p,d}^3 d = \Theta\left(\frac{\log^{3/2} n}{\sqrt{d}}\right)$. If $d \ll \log^3 n$, the exponent $\exp\left(\frac{\log^{3/2} n}{\sqrt{d}}\right)$ grows faster than any polylogarithmic function of n , which means that the obtained result is better than that of Lemma 3. Unfortunately, the upper bound of Theorem 7 is still $1/t_{p,d}^2$ times larger than the lower bound, although this margin is much smaller than the ‘main’ term $e^{t_{p,d}^3 d}$. This exponent still grows slower than any power of n , but we believe that for $d = \Theta(\log n)$, the number of triangles is linear (or almost linear) in n .

Numerical results

So far, all the presented results are strictly theoretical and refer to the asymptotic case. But real-life networks have a limited number of nodes, and it is not clear whether our theoretical results might be applied to their description. To verify this hypothesis, we conducted a few simple numerical experiments.

Our aim is to compare the clustering coefficient and the number of triangles in random geometric and Erdős–Rényi graphs. Here we use the global clustering coefficient (GCC), or transitivity, which is defined as follows:

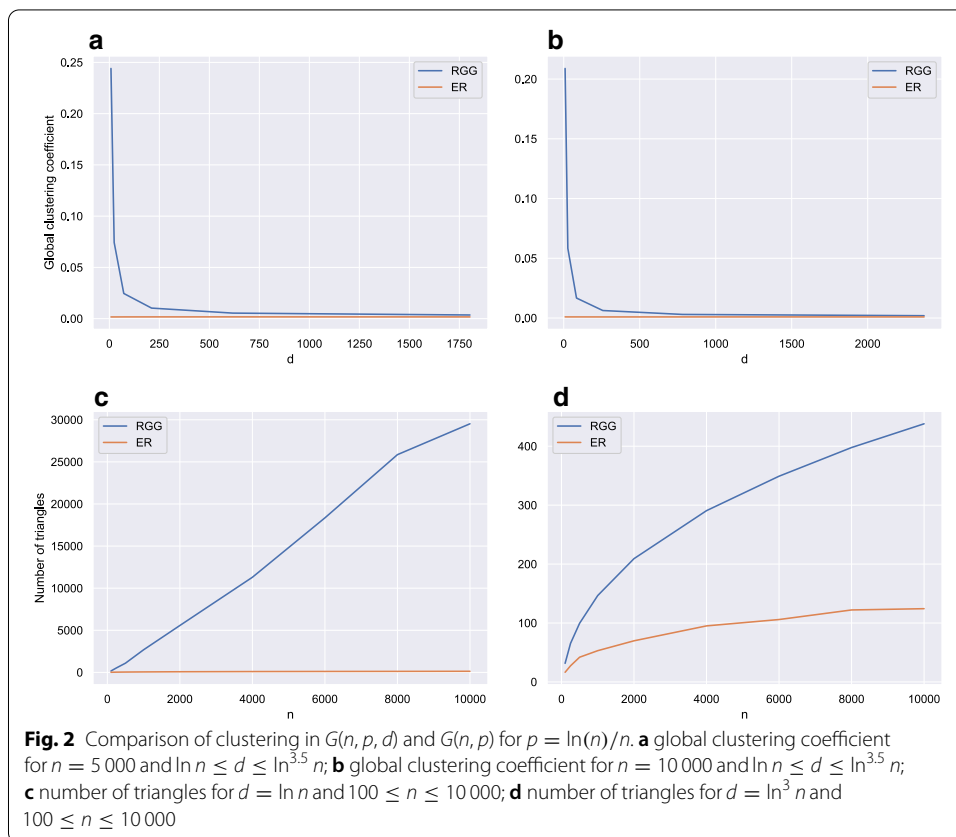
$$GCC = \frac{3 \times \text{number of triangles}}{\text{number of triplets}},$$

where a triplet is a configuration of three nodes connected by at least two edges. This quantity represents then the proportion of ‘closed’ triplets. In other words, the GCC is the probability that two of my ‘friends’ are also ‘friends’ to each other. In general, this is

a good quantitative expression of the clustering level in a network (see Wasserman and Faust 1994).

Figure 2 illustrates the difference between a random geometric graph and an Erdős–Rényi graph in terms of the GCC and the number of triangles. For our experiments, we took $p = \ln n/n$, a quite popular regime, usually called *significantly sparse*. Figure 2a, b show the average GCC (over 20 iterations) of $G(n, p, d)$ and $G(n, p)$ for $n = 5000$ and $n = 10000$, respectively, and for $\ln n \leq d \leq \ln^{3.5} n$. As expected, the difference is large when d is relatively small. As d increases, the difference goes to 0, and for $d = \ln^3 n$ (617 and 781, respectively), it equals 0.002. Since the GCC of $G(n, p)$ is simply p for large n , the GCC significantly higher than p gives a reason to suppose that the network has the underlying geometry with small d . On the other hand, the GCC close to p means that a low-dimensional graph representation is most likely impossible.

Unfortunately, our results and the result of Bubeck et al. (2016) do not give the exact value of constants for given n , and we cannot try them in practice. However, we can compare the growth rate of the number of triangles for different d . Fig. 2c, d show how fast the number of triangles grows with $n \rightarrow \infty$ for $d = \ln n$ and $d = \ln^3 n$, respectively. The number of triangles in $G(n, p)$, of course, does not depend on d and grows as $\ln^3 n$ on both pictures. As for $G(n, p, d)$, the number of triangles grows almost linearly in n for $d = \ln n$, while $d = \ln^3 n$ gives a ‘no geometry’ situation of the corresponding Erdős–Rényi graph, up to a multiplication factor.



To conclude, we see a possible extension of this theoretical work in obtaining practical results with explicit bounds on the expected number of triangles. Such bounds would help to determine whether a network has an underlying geometry. Moreover, if this is the case, an interesting problem is to determine (perhaps in real-life tasks) the dimension of the underlying geometric space. This would help, for instance, to make embedding of big data sets more efficient by better choice of the embedding dimension.

Conclusion

As we have seen, high-dimensional random geometric graphs in the sparse regime always fail to create really large communities. It would be natural to expect that these graphs do not differ in any way from Erdős–Rényi graphs; however, for $d \ll \log^3 n$, they show a rather high tendency for clustering (my ‘friends’ are connected with high probability). Is it true that in the opposite case $G(n, p, d)$ and $G(n, p)$ resemble each other? We believe in the conjecture stated in Bubeck et al. (2016), which proposes the positive answer to this question. Since a similar conjecture was proved in that work for the dense regime, the situation does not look hopeless. However, the technique applied in the dense regime cannot be easily extended to the sparse regime. Any result describing the total variation between $G(n, p, d)$ and $G(n, p)$ in this regime would be very interesting.

In the present paper and in Devroye et al. (2011), Bubeck et al. (2016), the case $d \geq \log n$ is always considered. What happens if d grows at a lower pace? What is the value of the clique number? We do not have the answers for this regime. But, obviously, the theoretical framework may differ quite a lot from what we used in our work.

As for triangles, it is not hard to prove that the number of triangles can be approximated by the Poisson distribution with an appropriate parameter that is, of course, the expected number of triangles. Hence, we need sharper bounds on this quantity, especially in the case $d \ll \log^3 n$. We are convinced that the upper bound in Theorem 7 cannot be improved, and the statement holds true for $\log n \ll d \ll \log^2 n$.

For sure, apart from the description of cliques and communities, many properties of high-dimensional random geometric graphs remain unexplored: connectivity, the existence of the giant component, the chromatic number, to name but a few. But even for fixed d , the results describing these properties require quite complex methods, so we do not expect immediate breakthroughs in this direction.

The previous section presents some numerical results that already might be useful for practical purposes. Let us discuss some possible further work in this direction. Firstly, we believe that the cliques (especially triangles) might be useful for community detection in networks with a geometric structure. Secondly, we think that some of the ideas introduced in this paper can help determine if a network has an underlying geometry. The latter is important because if it is known that the nodes are embedded in some space, then one can hope to make a lower-dimensional representation of the network structure or to use its geometric properties (e. g., two distant nodes cannot have a common neighbor). However, this requires more accurate bounds on the number of triangles with explicit constants. We did not pursue this goal and concentrated only on asymptotic results. Finally, the results obtained above can be helpful for the investigation of possible multiple correlations in data sets with a very large number of features.

Abbreviations

CDF: Cumulative distribution function; GCC: Global clustering coefficient.

Acknowledgements

Not applicable.

Authors' contributions

KA suggested the principal ideas and the topic of research. AB designed the research. KA and AB wrote, reviewed and revised the manuscript. Both authors read and approved the final manuscript.

Funding

The authors have no funding sources to acknowledge for this study.

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 8 April 2020 Accepted: 3 November 2020

Published online: 23 November 2020

References

- Albert R, Barabási A-L (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512. <https://doi.org/10.1126/science.286.5439.509>
- Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74(1):47–97. <https://doi.org/10.1103/RevModPhys.74.47>
- Alon N, Spencer J (2004) *The probabilistic method*. Wiley, New York
- Appel M, Russo R (2002) The connectivity of a graph on uniform points on $[0, 1]^d$. *Stat Probabil Lett* 60(4):351–357. [https://doi.org/10.1016/S0167-7152\(02\)00233-X](https://doi.org/10.1016/S0167-7152(02)00233-X)
- Arias-Castro E, Bubeck S, Lugosi G (2015) Detecting positive correlations in a multivariate sample. *Bernoulli* 21(1):209–241. <https://doi.org/10.3150/13-BEJ565>
- Avrachenkov K, Bobu A (2019) Cliques in high-dimensional random geometric graphs. In: Cherifi H, Gaito S, Mendes J, Moro E, Rocha L (eds) *Complex Networks and Their Applications VIII*. Complex networks 2019. Springer, Berlin, pp 591–600. https://doi.org/10.1007/978-3-030-36687-2_49
- Baccelli F, Błaszczyszyn B (2010) Stochastic geometry and wireless networks: Volume ii applications. *Found Trends Network* 4(1–2):1–312. <https://doi.org/10.1561/13000000026>
- Barthélemy M (2011) Spatial networks. *Phys Rep* 499(1–3):1–101. <https://doi.org/10.1016/j.physrep.2010.11.002>
- Blumenson L (1960) A derivation of n -dimensional spherical coordinates. *Am Math Mon* 67(1):63–66. <https://doi.org/10.2307/2308932>
- Bollobás B (2001) *Random graphs*. Cambridge University press, Cambridge, UK. <https://doi.org/10.1017/CBO9780511814068>
- Brieden A, Gritzmann P, Kannan R, Klee V, Lovász L, Simonovits M (2001) Deterministic and randomized polynomial-time approximation of radii. *Mathematika* 48(1–2):63–105. <https://doi.org/10.1112/S0025579300014364>
- Bubeck S, Ding J, Eldan R, Rácz M (2016) Testing for high-dimensional geometry in random graphs. *Rand Struct Algorith* 49(3):503–532. <https://doi.org/10.1002/rsa.20633>
- Devroye L, Györfy A, Lugosi G, Udina F (2011) High-dimensional random geometric graphs and their clique number. *Electron J Probabil* 16:2481–2508. <https://doi.org/10.1214/EJP.v16-967>
- Erdős P, Rényi A (1960) On the evolution of random graphs. *Publ Math Inst Hung Acad Sci Ser A* 5(1):17–60
- Fronczak A, Fronczak P, Holyst J (2003) Mean-field theory for clustering coefficients in barabási-albert networks. *Phys Rev E* 68(4):046126. <https://doi.org/10.1103/PhysRevE.68.046126>
- Gilbert E (1961) Random plane networks. *J Soc Ind Appl Math* 13:266–267
- Haenggi M, Andrews J, Baccelli F, Dousse O, Franceschetti M (2009) Stochastic geometry and random graphs for the analysis and design of wireless networks. *IEEE J Sel Areas Commun* 27(7):1029–1046. <https://doi.org/10.1109/JSAC.2009.090902>
- Higham D, Rašajski M, Pržulj N (2008) Fitting a geometric graph to a protein-protein interaction network. *Bioinformatics* 24(8):1093–1099. <https://doi.org/10.1093/bioinformatics/btn079>
- Krioukov D, Papadopoulos F, Kitsak M, Vahdat A, Boguńá M (2010) Hyperbolic geometry of complex networks. *Phys Rev E* 82(3):036106. <https://doi.org/10.1103/PhysRevE.82.036106>
- Lee Y, Kim W (2014) Concise formulas for the surface area of the intersection of two hyperspherical caps. KAIST Technical Report
- McDiarmid C (2003) Random channel assignment in the plane. *Rand Struct Algorithms* 22(2):187–212. <https://doi.org/10.1002/rsa.10077>
- McDiarmid C, Müller T (2011) On the chromatic number of random geometric graphs. *Combinatorica* 31(4):423–488. <https://doi.org/10.1007/s00493-011-2403-3>
- McPherson M (2004) A blau space primer: prolegomenon to an ecology of affiliation. *Ind Corp Change* 13(1):263–280. <https://doi.org/10.1093/icc/13.1.263>
- Müller T (2008) Two-point concentration in random geometric graphs. *Combinatorica* 28(5):529. <https://doi.org/10.1007/s00493-008-2283-3>

- Nekovee M (2007) Worm epidemics in wireless ad hoc networks. *New J Phys* 9(6):189. <https://doi.org/10.1088/1367-2630/9/6/189>
- Penrose M (1999) On k -connectivity for a geometric random graph. *Rand Struct Algorithms* 15(2):145–164
- Penrose M (2003) *Random geometric graphs*, vol 5. Oxford University press, Oxford. <https://doi.org/10.1093/acprof:oso/9780198506263.001.0001>
- Pottie G, Kaiser W (2000) Wireless integrated network sensors. *Commun ACM* 43(5):51–58. <https://doi.org/10.1145/332833.332838>
- Preciado V, Jadbabaie A (2009) Spectral analysis of virus spreading in random geometric networks. In: *Proceedings of the 48th IEEE Conference on Decision and Control (CDC) Held Jointly with 2009 28th Chinese Control Conference*, pp. 4802–4807. <https://doi.org/10.1109/CDC.2009.5400615>. IEEE
- Wasserman S, Faust K (1994) *Social network analysis: methods and applications*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511815478>
- Xiao H, Yeh E (2011) Cascading link failure in the power grid: A percolation-based analysis. In: *2011 IEEE International Conference on Communications Workshops (ICC)*, pp. 1–6. <https://doi.org/10.1109/iccw.2011.5963573>. IEEE

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
