Testing for high-dimensional geometry in random graphs

Ronen Eldan
Weizmann Institute of Science
ronen.eldan@weizmann.ac.il

Dan Mikulincer
Weizmann Institute of Science
dan.mikulincer@weizmann.ac.il

Overview

How to distinguish between distrbutions on random graphs?
We study the total variation between

Erdős-Rényi random graphs

and

Random geometric graphs

Where the underlying metric space
Is taken to be a d-dimensional Ellipsoid.

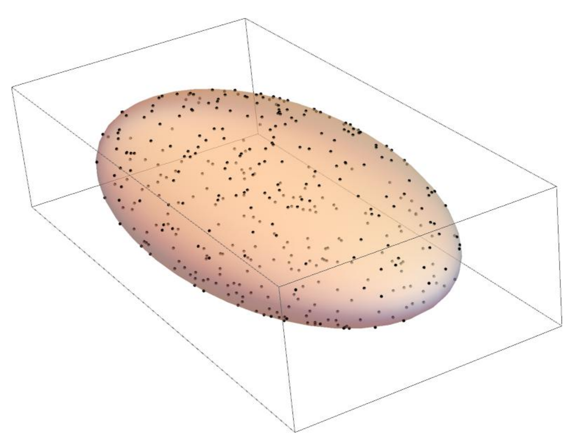
Definitions

Let $\alpha \in \mathbb{R}^d$, denote by E_α a d-dimensional elliposid whose axes correspond to the entries of α.
For a given $p \in (0,1)$ set $t_{p,\alpha}$ to be such that when X, Y are chosen randomly and uniformly from E_α it holds that

$$\mathbb{P}(\langle X, Y\rangle \geq t_{p,\alpha}) = p$$

Random graphs

Fix $p \in (0,1)$, we denote by $G(n,p)$ the standard Erdős-Rényi distribution on n vertices.



For a given $\alpha \in \mathbb{R}^d$ we define $G(n,p,\alpha)$ to be a random geometric graph with vertices sampled uniformly from E_α. Two vertices v, u are connected by an edge if and only if $\langle v, u\rangle \geq t_{p,\alpha}$.

We are interested in studying the total variation between these two models, denoted by $\mathrm{TV}(G(n,p), G(n,p,\alpha))$.

As we let n and α vary, we are intersted in the question: What conditions must we require from α compared to n, so that $\mathrm{TV}(G(n,p), G(n,p,\alpha))$ remains bounded away from 0?.

Main result

For a $\alpha \in \mathbb{R}^d$ and $q > 1$, let

$$||\alpha||_q = \left(\sum_{i=1}^{d} \alpha_i^q\right)^{\frac{1}{q}},$$

denote the L_q norm of α.

Theorem. (a) Let $p \in (0,1)$ be fixed and assume $\left(\frac{||\alpha||_2}{||\alpha||_3}\right)^6 / n^3 \to 0$. Then

$$\mathrm{TV}(G(n,p), G(n,p,\alpha)) \to 1.$$

(b) Furthermore, if $\left(\frac{||\alpha||_2}{||\alpha||_4}\right)^4 / n^3 \to \infty$ then,

$$\mathrm{TV}(G(n,p), G(n,p,\alpha)) \to 0.$$

Comments:

- There is a gap between the bounds (a) and (b).
- In the isotropic case (when $\alpha_i = 1$ for all i) the gap disappears and both quantities reduce to d.

Proof ideas

The bound (a) is proven using a variant of counting triangles. Indeed, the triangle inequality suggests that $G(n,p,\alpha)$ should have more triangles than $G(n,p)$.
However, the number of triangles in a random graph has unnecessarily high variance. Thus, instead of counting triangles, we count *signed triangles* which offer a decrease in variance. If A is the adjacency matrix of G we define the number of signed triangles in G to be:

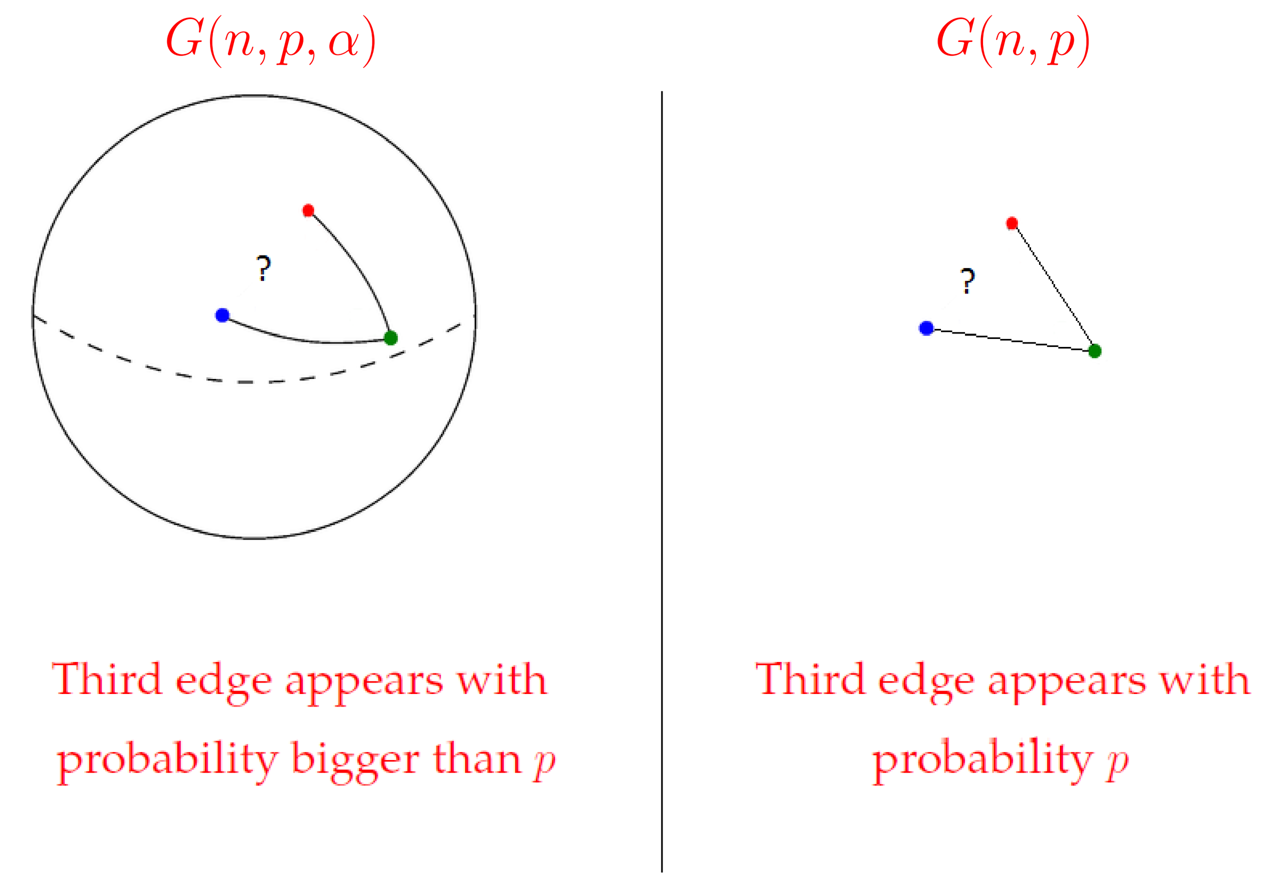
$$\tau(G) = \mathrm{Tr}((A-p)^3).$$

Bound (b) is a result of a multi-dimensional version of the Entropic Central Limit Theorem. Using Pinsker's inequality we may bound the total variation with the relative entropy

$$\mathrm{TV}(G(n,p), G(n,p,\alpha)) \leq \sqrt{\frac{1}{2}\mathrm{Ent}[G(n,p)||G(n,p,\alpha]}$$

It is then enough to consider the graphs as measurements of continous matrices. Where $G(n,p)$ corresponds to a Gaussian matrix, and $G(n,p,\alpha)$ to a sum of Wishart matrices.

Intuition



Future directions

- **Closing the gap.** It stands to reason that the use of the entropic CLT does not give a tight characterization.
- **Signed Triangles.** Show that signed triangles are an optimal statistic in this setting.
- **Added Randomness.** Study what happens when there is an added probability for edges not to exist in $G(n,p,\alpha)$.

References

[1] S. Bubeck, J. Ding, R. Eldan, M.Z. Rácz. Testing for high-dimensional geometry in random graphs. Preprint at http://arxiv.org/abs/1411.5713.

[2] S. Bubeck, S. Ganguly. Entropic CLT and phase transition in high-dimensional Wishart matrices. Preprint at http://arxiv.org/abs/1509.03258.