Fractal Gaussian Networks: A sparse random graph model based on Gaussian Multiplicative Chaos

Subhroshekhar Ghosh [1] **Krishnakumar Balasubramanian** [2] **Xiaochuan Yang** [3]

Abstract

We propose a novel stochastic network model, called Fractal Gaussian Network (FGN), that embodies well-defined and analytically tractable fractal structures. Such fractal structures have been empirically observed in diverse applications. FGNs interpolate continuously between the popular *purely random* geometric graphs (a.k.a. the Poisson Boolean network), and random graphs with increasingly fractal behavior. In fact, they form a parametric family of *sparse* random geometric graphs that are parametrized by a fractality parameter ν which governs the strength of the fractal structure. FGNs are driven by the latent spatial geometry of Gaussian Multiplicative Chaos (GMC), a canonical model of fractality in its own right. We explore the natural question of detecting the presence of fractality and the problem of parameter estimation based on observed network data. Finally, we explore fractality in community structures by unveiling a natural stochastic block model in the setting of FGNs.

1. Stochastic Networks and Fractality

The *unreasonable effectiveness* of stochastic networks. Stochastic networks have emerged as one of the fundamental modelling paradigms in the last few decades in our efforts to effectively understand the structures underlying vast amounts of data with increasing complexity, in order to capture the effects of latent factors and their mutual interactions. At a broad level of abstraction, this involves *nodes* representing agents, and *edges* (weighted or otherwise) that embody the interactions between these agents. Indeed, the ubiquity of statistical network models in the modern applied sciences may justifiably remind one of the famous article of E.P. Wigner on the *unreasonable effectiveness* of mathematics in the natural sciences (c.f. (Wigner, 1990)).

Within the domain of statistical networks, many popular modelling formulations have been proposed and investigated, in order to understand different types of phenomena in large complex systems. These include the fundamental Erdős-Rényi random graph model, the preferential attachment model and its variants, random geometric graphs, graphons, the stochastic block model and its various avatars, small world networks like the Watts-Strogatz model, models of scale-free networks, to provide a partial list of examples (see, e.g., (Albert & Barabási, 2002),(Strogatz, 2001), (Lovász, 2012), (Erdös et al., 1959), (Penrose et al., 2003), (Holland et al., 1983), (Orbanz & Roy, 2014)). As a preview to connect our present contribution to this classical literature, here we aim to propose a novel paradigm of statistical networks with a view to capturing fractal phenomena.

The application domains for stochastic network models are diverse, ranging from the world-wide web and inter/intranets, collaboration networks in academia, and social and communication networks. Indeed, modern day network science has developed into a unique discipline of its own, for an overview of which we refer the reader to any amongst a multitude of excellent texts - at this point we mention (Barabási et al., 2016), (Lewis, 2011), (Mezard et al., 2009), (Crane, 2018), (Watts, 2004), (Bickel & Chen, 2009), (Chung et al., 2006) , (Bollobás et al., 2010), (Van Der Hofstad, 2016), (Spielman, 2010), (Caldarelli, 2007), (Jackson, 2010), only to provide a partial list.

Fractal structures in large scale networks. An important feature which has come to the fore in recent investigations of networks is the emergence of inherent fractal structures in diverse application domains. Heuristically, fractal structures are often characterized by non-standard and anomalous behavior of various scaling and growth exponents, and (truncated) power law tails for naturally associated statistics (c.f., (Falconer, 2004), (Mandelbrot, 1983), (Avnir et al., 1998)).

There are many instances of emergence of fractality in networks. To provide a detailed example, in human mobility

[1]Department of Mathematics, National Uuniversity of Singapore. [2]Department of Statistics, University of California, Davis. [3]Department of Mathematics, University of Luxemborg. Correspondence to: Subhroshekhar Ghosh <subhrowork@gmail.com>, Krishnakumar Balasubramanian <kbala@ucdavis.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

networks, it has been observed that the layout of the way-points in the trajectories and the boundaries of popular sojourn domains exhibit fractal properties on a global scale, and the flight/pause times and inter-contact times between the agents exhibit power law tails (see, e.g., (Lee et al., 2009), (Lee et al., 2011), (Rhee et al., 2011)). Another important class of examples is the discovery of fractal structures in transportation networks, like urban bus transport networks and railway networks ((Benguigui, 1992), (Pavón-Domínguez et al., 2017), (Murcio et al., 2015), (Salingaros, 2003)) and drainage networks ((Rinaldo et al., 1992), (Rinaldo et al., 1993), (La Barbera & Rosso, 1989) (Claps et al., 1996)).

Fractality and multifractality are also known to arise in the context of scale-free and other complex networks ((Song et al., 2005), (Song et al., 2006), (Kim et al., 2007)), internet traffic ((Caldarelli et al., 2000)) and financial networks; in fact, financial data in general present an important class of problems where fractal properties are known to occur (c.f., (Caldarelli et al., 2004), (de la Torre et al., 2017), (Mandelbrot, 2013), (Mandelbrot & Hudson, 2010), (Inaoka et al., 2004), (Evertsz, 1995)). Fractal phenomena have emerged in sociological and ecological networks, dense graphs and graphons ((De Florio et al., 2013), (Hill et al., 2008), (Gao et al., 2012), (Palla et al., 2010), (Lyudmyla et al., 2017)), biological neural networks ((Bassett et al., 2006)), network dynamics ((Orbach, 1986), (Goh et al., 2006)) and even in the field of development economics ((Barrett & Swallow, 2006)).

Towards a parametric model of fractality in sparse networks: the Fractal Gaussian Network model (FGN). In view of the diversity of settings in which fractality has been observed to occur in networks, it is natural to investigate concrete mathematical models of fractality in networks which, on one hand, are amenable to rigorous theoretical analysis, and on the other hand, allow a broad enough horizon to study a reasonably wide class of phenomena. Furthermore, it would be of great interest to have a parametric statistical model, e.g. in the spirit of exponential families of classical parametric statistics ((Bickel & Doksum, 2015)). This will open up a natural programme of investigation in terms of parameter estimation, tests of hypothesis with regard to fractal structures and examination of the model under parametric modulation.

In this work, we propose a statistical model for network data that aims to understand such fractal structures in a rigorous and analytically tractable manner. Based on a latent random field structure accorded by *Gaussian Multiplicative Chaos* (GMC), a canonical model of fractal phenomena in various branches of natural and applied sciences, we call it the *Fractal Gaussian Network* model, which we will abbreviate henceforth as FGN.

2. Gaussian Multiplicative Chaos : an overview

Gaussian Multiplicative Chaos (GMC) form a natural family of random fractal measures. Roughly speaking, the GMC is defined on a Euclidean base space (e.g., a domain $\Omega \in \mathbb{R}^d$, scaled to have volume 1), and originates from an underlying centered Gaussian field $(X(x))_{x\in\Omega}$. Typically, on Euclidean spaces the Gaussian field X is taken to be translation invariant and *logarithmically correlated*. This entails, for example, that the covariance kernel K of the Gaussian field X has the following form :

$$K(x,y) = \ln_+ \frac{T}{|x-y|} + g(x-y),$$

where $T > 0$ and g is a bounded continuous function. Such fields arise naturally in many areas of mathematics, statistical physics and their applications, an important example being the celebrated Gaussian Free Field model (see, e.g., (Sheffield, 2007) and the references therein).

For any $\gamma > 0$ (with $\gamma^2 < 2d$ in order to ensure non-degeneracy of the limiting measure) and a Radon measure μ on Ω, we consider the random measure defined on Ω that is given, heuristically speaking, by the formula

$$\mathrm{d}M^\gamma(x) := \exp(\gamma X(x) - \frac{\gamma^2}{2}\mathbb{E}[X(x)^2])\mathrm{d}\mu(x). \quad (1)$$

In the common setting of translation-invariance and μ the d-dimensional Lebesgue measure, this simply reduces to the form $\mathrm{d}M^\gamma(x) = C_\gamma \exp(\gamma X(x))\mathrm{d}x$, which is the setting on which we are going to focus in this article. It is known that in this case the expected measure $\mathbb{E}[\mathrm{d}M^\gamma(x)] = \mathrm{d}x$, i.e. the Lebesgue measure, which provides a convenient background measure to compare a typical realization of the GMC with. We will provide a more technical discussion of GMC in Section 10, and refer the reader interested in a full treatment to the excellent survey (Rhodes et al., 2014) and the references contained therein.

GMC is a canonical model of fractal behavior in nature, endowed with statistical invariance properties that make it both an attractive mathematical structure as well as a robust modelling. Originating in the study of quantum field theory ((Høegh-Krohn, 1971), (Simon, 2015)) and the seminal work of J.P. Kahane ((Kahane, 1985), (Kahane & Peyriere, 1976)), it has many applications to fundamental problems like the study of quantum gravity (see, e.g., (Duplantier & Sheffield, 2009), (Duplantier & Sheffield, 2011)), as well as applied sciences where the GMC and related ideas have been effectively used to model volatility in financial assets and problems of turbulence (see, e.g., (Liu et al., 1999), (Duchon et al., 2012), (Kolmogorov, 1941), (Kolmogorov, 1962), (Fyodorov et al., 2010) and the surrounding literature).

A crucial point is that, because of the logarithmic singularity of the covariance kernel, the Gaussian field X is usually not well-defined as a function, but can be made sense of only as a Schwarz distribution (that acts on a smooth enough class of functions). Consequently, the equation (1) (that essentially purports to give a formulaic description of the GMC in terms of a random density with respect to a Radon measure) is only valid as a heuristic description. In fact, significant technical effort needs to be dedicated to make rigorous sense of the GMC as a random measure (without a well-defined density), a natural path to which is via approximating Gaussian fields for which everything is well-defined and taking limits.

The fact that the density in (1) does not exist as a well-defined, albeit random, function indicates that as a random measure GMC is indeed almost surely a *fractal measure*. This can also be demonstrated rigorously, and it can be shown that the GMC a.s. has a fractal dimension $d - \frac{\gamma^2}{2}$. It may be noted that, compared to the *ambient dimension* d, it is this fractal dimension that is more intrinsic to the GMC measure.

3. Generating Fractal Gaussian Networks

We next proceed to describe the construction of the FGN based on the GMC. To this end, we will require the following ingredients :

- An integer $d > 0$, a parameter $\gamma > 0$ with $\gamma^2 < 2d$, and a domain $\Omega \subset \mathbb{R}^d$ with $\mathrm{Vol}(\Omega) = 1$.
- A centered Gaussian random field X that lives on Ω, with a logarithmically singular covariance kernel at the diagonal.
- A realization of the GMC M^γ on the domain Ω and based on the random field X .
- A *size parameter* n, which is a positive integer (to be thought of as large but finite).
- A *connectivity threshold* σ (whose natural size will turn out to be $\propto n^{-1/d}$)
- A Poisson random variable N that is distributed with mean $nM^\gamma(\Omega)$

With the above ingredients in hand, we now proceed to construct the FGN model via the following steps:

- Sample N-many points, denoted by $V := \{x_1, \ldots, x_N\}$ at random from the given realisation M^γ of the GMC measure (after normalizing it to have total mass 1). The points in V will form the nodes of the FGN.
- Connect each x_i with any other x_j that is within distance σ of x_i. It turns out that there are multiple ways of implementing such connectivity that, broadly speaking, leads to similar behavior of various network statistics.
  - A direct approach to just connect two points in V if and only if they are within distance σ of each other.
  - A refined approach to connect two vertices $x_i, x_j \in V$ with probability $\propto \exp(-\frac{\|x_i - x_j\|^2}{\sigma^2})$. This allows for the possibility of long range connectivity.

In the last step of constructing the edges, it is the latter, more refined approach of adding edges randomly according to a Gaussian kernel that we will follow for the rest of this paper. However, we note in the passing that we believe the key phenomena will largely be true for the direct approach of connecting vertices merely based on their Euclidean separation. It turns out that $\mathbb{E}[M^\gamma(\Omega)] = |\Omega| = 1$, therefore $\mathbb{E}[N] = n\mathbb{E}[M^\gamma(\Omega)] = n$, so n is the natural large parameter indexing a growing network size.

Single pass and multi pass observation models. Our data access model is that we have access to the *combinatorial data* of the graph. In other words, our information will consist merely of a graph with vertices labelled $\{1, \ldots, N\}$ and vertices i and j connected by an edge if and only if the points x_i and x_j are connected in the above geometric graph. Thus, the spatial geometric structure of the GMC is purely a *latent factor* in the FGN, which we have no direct access to in our statistical investigations.

We will explore two different observation models for the FGN. One observation model, which we call the *single pass observation model* is that we have access to a single realization of the network, in the regime where the network size parameter n is very large. The other observation model, which we call the *multi pass observation model*, entails that we have access to a moderately large number m of i.i.d. copies of the network, in the regime where the size parameter n is also moderately large.

Both these observation models are well-motivated as modelling paradigms. In particular, for the FGN model, it may be noted that the underlying Gaussian field $X(x)$ is often taken to be translation invariant on $\mathbb{R}^d$. Hence, if two samples of the spatial geometric graph are obtained from two sub-domains of the full space that are translates of each other (i.e., we observe the nodes and edges for points in two domains $\mathcal{D}$ and $\mathcal{D} + x_0$ for some vector $x_0 \in \mathbb{R}^d$), then the subgraphs so obtained are identically distributed (because of the translation invariance of the underlying Gaussian field). On the other hand, if the sub-domains are well-separated in the ambient space, then they can be taken to be approxi-

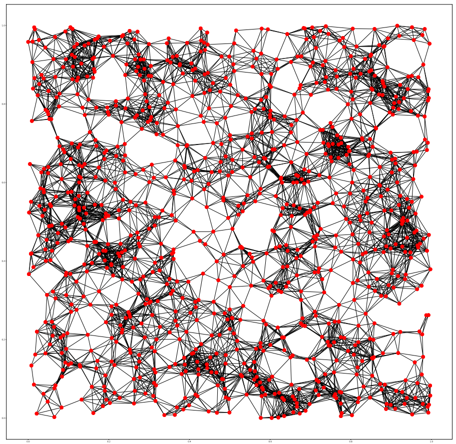

Figure 1. A realization of the FGN.

mately independent because of decay of correlations of the Gaussian field.

Thus, several approximately independent and identically distributed realizations of the same FGN can be obtained by taking samples of a very large, *universal* network based on surveying spatially similar and well-separated regions. Since the fractal properties may be reasonably assumed to be similar in different segments of a very large network, this provides us with a way of obtaining multiple samples from a FGN model that can capture fractal structures similar to the original graph. This can be compared, for example, with taking localized snapshots of a different parts of a vast communication network like the internet.

4. Properties of the FGN

Inherent fractal structure of the FGN. The inherent fractal nature of a typical realization of the GMC measure induces fractality in the FGN. For instance, one consequence of fractality in terms of the network structure is a large measure of heterogeneity, often manifested in terms of the irregular distribution of nodes in the form of dense clusters and rarefied neighborhoods in the graph.

The GMC is characterized by regions of high concentration of measure, interspersed with regions of low mass distribution. To see this in more detail, we refer the reader to Figure 1 in (Rhodes et al., 2014). In fact, a progressive increase in the irregularity of the GMC can be observed as the parameter γ increases. The FGN, because of its latent spatial geometry being derived from the GMC, also inherits these heterogeneities in its graphical structure, characterized by certain vertex clusters of high connectivity interspersed with sparsely connected vertices (see, e.g., Figure 1 for an illustration), with such heterogeneous effects increasing in intensity as the parameter γ increases in value.

The latent geometry of the *social space*. It may be observed that, once the realization of the GMC measure is in our hands, the rest of the construction of the FGN is spatial geometric in nature, and can actually be carried out for any non-negative measure on the domain Ω - random or otherwise.This spatial geometric construction employs the commonly used technique for the construction of random geometric graphs (RGG, c.f. (Penrose et al., 2003), (Gilbert, 1961)), popularly considered in the setting of the uniform distribution on Ω (which is going to be our "pure noise" case and the point of comparison with the FGN regarding the presence of fractal structures).

At this point, a word is in order regarding the spatiality inherent in the construction of the FGN. It turns out that many natural applications of stochastic networks have spatiality built into their construction - mobility networks, transportation networks or drainage networks are all examples of this phenomenon. But even more generally, our construction of the FGN does not necessitate the ambient Euclidean space $\mathbb{R}^d$ to correspond to our application in a *physical sense*. In fact, the ambient space $\mathbb{R}^d$ can be taken to be the *feature space* obtained from a feature mapping of the nodes, whose specifics can be completely problem-dependent. This is exemplified by its applications in the social networks, where the feature mapping corresponding to a person corresponds to his/her interests, and two persons are connected in the social network if their interests (i.e., feature vectors) are close in the metric of the latent *social space* (c.f., (Jackson, 2010), (Rácz et al., 2017), (Sarkar & Moore, 2006), (Grover & Leskovec, 2016)).

Such graphs are of interest as statistical networks in both low and high dimensional spatial settings (see, e.g., (Bubeck et al., 2016), (Bubeck et al., 2015), (Mossel et al., 2018), (Bubeck et al., 2017), (Bubeck & Ganguly, 2018), (Rácz & Richey, 2019)). *Physical spatiality* would often correspond to a *low* ambient dimension (as in the case of transportation or drainage networks), whereas *latent spatiality* in the social/feature space may naturally correspond to a *relatively high* ambient dimension d. It may be pointed out that the FGN model encompasses both low and high dimensions of the latent space, thereby catering to both types of spatial structure.

The connectivity threshold σ, locality and the sparse regime. We now undertake a brief discussion of the choice of the threshold σ. We will study the FGN in the regime in which it is a *sparse graph*, which entails that a given node will typically have $O(1)$ neighbors. This is most natural in the context of most real world networks - even though

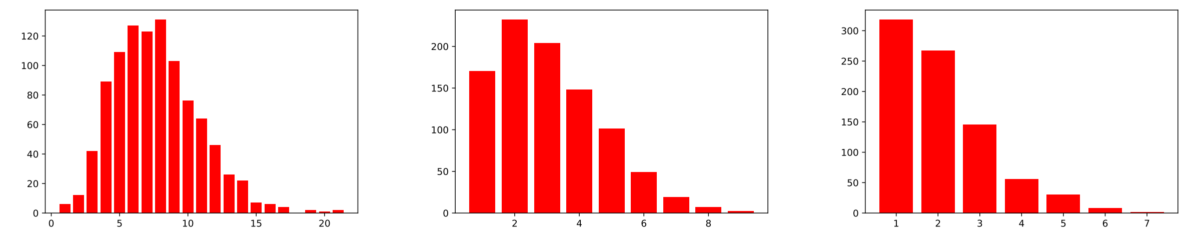

Figure 2. Degree distribution of non-isolated nodes of the FGN: the value of ν increases from left to right.

the total network might be huge and highly complex, seen from the viewpoint of a particular node it has a *finite local neighborhood*, which does not scale with the growing size of the network (see, e.g., (Johnson, 1977), (Krzakala et al., 2013), (Guédon & Vershynin, 2016), (Batagelj & Mrvar, 2001) and the references therein).

It will turn out from our detailed analysis of the FGN is Section 11 that, for any value of γ, the normalization $\sigma = \frac{1}{\sqrt{\pi}}\rho^{1/d}n^{-1/d}$ will lead to, in expectation, ρ neighbors for a given point under our connection model. We will call ρ the *density* or the *degree* parameter. From a statistical point of view, the density parameter ρ can be learnt by looking at the degrees of vertices (possibly sampled at several disjoint neighborhoods of the graph), rendering the density ρ a *local parameter* in the FGN model. Local parameters are much easier to investigate because they can be learnt by sampling small local neighborhoods, which for practical purposes can be taken to be approximately independent if they are well separated (e.g., in the graph distance).

On the other hand, in real world networks, fractality is often observed at the scale where one *zooms out*, i.e., at mesoscopic scales or higher (c.f., (Franović & Miljković, 2009), (Pook & Janßen, 1991), (Daqing et al., 2011)). This necessitates the investigation of fractality to be contingent on more global aspects of the FGN, which makes it much more challenging but at the same time more interesting to study and is the principal focus of this article.

The intrinsic *fractality parameter* ν. For the FGN model, the key determinant of fundamental network statistics turns out be the quantity $\nu = \frac{\gamma^2}{d}$, which we refer to as the *fractality parameter*. Accordingly, we will maintain a particular consideration for the *fractality parameter* ν in our statistical analysis of the FGN model.

5. FGN as a parametric statistical model

Interpolating homogeneity and fractality. We will formulate our analysis of the FGN as a statistical model of network data in terms of the quantity ν, a choice that is well-motivated by the discussions in the preceding sections. It may be noted that, when $\nu = 0$ (equivalently, $\gamma = 0$), the GMC reduces to the Lebesgue measure, and we have a usual Poisson random geometric graph, which we will consider as the *pure noise* case in our setting. We will compare this against the presence of fractality in the network, a situation which would correspond to $\nu > 0$.

Thus, the FGN model *interpolates continuously* between Poisson random geometric graphs and networks with increasing degree of fractality, as the value of the parameter ν increases from 0. On a related note, it would also be of interest to learn the value of fractality parameter ν in its own right, which would correspond naturally to the problem of parameter estimation in the FGN model.

In our investigation of the standard statistical questions on the FGN model, such parameter estimation and testing (undertaken in Sections 5, 5), we will make extensive use of the statistics of small subgraph counts (in particular the edge counts). This is well-motivated by the effectiveness of small subgraph counts as statistical observables in the study of usual spatial network models (see, e.g., (Rácz et al., 2017), (Bubeck et al., 2016), and the references therein).

Degree distribution: Interpolating Poisson and power laws. We investigate the degree distribution of the FGN model empirically (c.f. Figure 2). We observe that, for small ν, the degree distribution is Poissonian, whereas with increasing values of the parameter ν, it deforms into a truncated power law like distribution. It may be noted that power laws and truncated power laws are ubiquitous in many real-world networks (c.f. (Albert & Barabási, 2002) and the references therein), whereas Poissonian behavior is a hallmark of classical *mean-field* models (like the Erdős-Rényi random graphs, c.f. (Van Der Hofstad, 2016)). As parametric statistical model, the FGN continuously interpolates between these two very different worlds which are the two major paradigms the distributions of degrees in networks.

Inferring the size parameter n. It may be noted that the parameter n driving the network size is not given to us in the combinatorial data that we can access. In some sense, it is also a latent parameter of the model that we do not directly focus on in our study of the fractal properties of the network.

However, statistical procedures typically utilize large sample effects, and for that purpose, it becomes imperative to develop an idea of the underlying size parameter n from the combinatorial graph.

To this end, we observe that, conditioned on the GMC, the network size N is a Poisson random variable with mean $nM^\gamma(\Omega)$. As such, if $\{Y_i\}_{i\geq 0}$ are i.i.d. Poisson random variables with mean $M^\gamma(\Omega)$, then $N = \sum_{i=1}^n Y_i$ in distribution. Consequently, for a given realization of the GMC, the quantity $\frac{N}{n} = \frac{1}{n}\sum_{i=1}^n Y_i \to M^\gamma(\Omega)$ a.s. as $n \to \infty$. As a result, $\frac{N}{n} = O(1)$ with high probability, as the size parameter $n \to \infty$.

Therefore, for the single pass observation model, in the regime of large size parameter n (which is the regime in which we envisage the FGN model), we may justifiably employ the network size N as an estimator for the latent size parameter n. In particular, on the logarithmic scale we may deduce that

$$\log N = \log n(1 + o_P(1)), \tag{2}$$

which is a form that will be particularly useful in our later analysis.

In the multi pass observation model, where we have m i.i.d. realizations of the FGN with node counts $N_1, N_2, \ldots, N_m$, we will use $\overline{N} = \frac{1}{m}\sum_{i=1}^m N_i$, which will strongly concentrate around its expectation n.

Estimating the fractality parameter ν. In this section we propose an estimator for the crucial fractality parameter ν in the FGN model. To this end, we will focus on small subgraph counts in the network, and utilize our analysis of edge counts (for details, see Section 12) in order to detect fractal structures. In this section, we will work in the setting $\gamma^2 < d$, so that the results of Section 12 would be applicable. Interestingly, this would correspond to the so-called L^2 *regime* in the theory of GMC, where many of the mathematical technicalities are known to be relatively more tractable.

In the single pass observation model, we consider the statistic

$$\hat{\nu}_{\mathrm{single}} := \frac{\log \mathcal{E}}{\log N} - 1 \tag{3}$$

as an estimator for ν, where $\mathcal{E}$ is the edge count of the FGN. To see why $\hat{\nu}_{\mathrm{single}}$ is a good estimator, we write

$$\begin{aligned}\log \mathcal{E} &= \log\left(\frac{\mathcal{E}}{n^{1+\nu}} \cdot n^{1+\nu}\right)\\ &= (1+\nu)\log n + \log(\mathcal{E}/n^{1+\nu})\\ &= \log n\left[1 + \nu + \frac{\log(\mathcal{E}/n^{1+\nu})}{\log n}\right].\end{aligned}$$

But Theorem 12.1 suggests that $\mathcal{E}/n^{1+\nu}$ is an O(1) quantity, which indicates that $\frac{\log \mathcal{E}}{\log n} \sim 1 + \nu$ as $n \to \infty$. But we may now make use of the fact that we have $\log N = \log n(1 + o(1))$ (as in (2)), which, coupled with the last equation, implies that $\frac{\log \mathcal{E}}{\log N}$ is approximately $1+\nu$, or equivalently $\frac{\log \mathcal{E}}{\log N} - 1$ is approximately ν in the regime of large size parameter n, as desired.

In the multi pass observation model, we have m i.i.d. samples of the FGN with $\mathcal{E}_i$ and N_i being the edge count and the vertex count of the i-th sample. Then we may define $\overline{\mathcal{E}}$ as the mean edge count $\overline{\mathcal{E}} := \frac{1}{m}\sum_{i=1}^m \mathcal{E}_i$ and $\overline{N}$ as the mean vertex count $\overline{N} := \frac{1}{m}\sum_{i=1}^m N_i$. We observe that, in the regime of large m, the mean edge count $\overline{\mathcal{E}}$ and the mean vertex count $\overline{N}$ strongly concentrate around their expectations. As such, in the regime of large m we have $\overline{\mathcal{E}} = \mathbb{E}[\mathcal{E}](1 + o_P(1))$ and $\overline{N} = \mathbb{E}[N](1 + o_P(1)) = n(1 + o_P(1))$.

This, coupled with our analysis of the single pass setting, naturally suggests consideration of the following estimator of ν in the multi pass observation model :

$$\hat{\nu}_{\mathrm{multi}} := \frac{\log \overline{\mathcal{E}}}{\log \overline{N}} - 1. \tag{4}$$

The efficacy of $\hat{\nu}_{\mathrm{multi}}$ as an estimator for ν follows from the afore-mentioned asymptotics of $\overline{\mathcal{E}}$ and $\overline{N}$ in the large m regime, which lead us to deduce that

$$\begin{aligned}\hat{\nu}_{\mathrm{multi}} =& \frac{\log \overline{\mathcal{E}}}{\log \overline{N}} - 1\\ =& \frac{(\log \mathbb{E}[\mathcal{E}] + \log(1 + o_P(1)))}{(\log \mathbb{E}[N] + \log(1 + o_P(1))) - 1}\\ =& (1+\nu) + o_P(1) - 1\\ =& \nu + o_P(1),\end{aligned}$$

where, in the last step, we have used the asymptotics of $\mathbb{E}[\mathcal{E}]$ in Theorem 12.1, coupled with a small parameter expansion of $\log(1 + x)$.

The estimator $\hat{\nu}$ has the form of a log-log plot between network observables and system size. Such log-log plots have been used effectively in studying growth exponents and fractal behavior in the phenomenological literature (cite), and therefore are well-motivated and thoroughly contextualized in the setting of fractal networks.

Detecting the presence of fractality. We examine the presence of fractality in the network by examining whether the combinatorial data of the graph points to the occurrence of such structures. As argued earlier, in the context of FGN this would entail determining whether $\nu = 0$ (absence of fractality), and compare it with the alternative possibility $\nu \geq \nu_0$ for some given threshold ν_0 (presence of a substantive degree of fractality). Choosing a positive threshold for the alternative, separated from 0, is a natural framework, because as discussed earlier the FGN interpolates *continuously* between homogeneity and gradually increasing fractality.

In this section, we will once again work in the setting $\gamma^2 < d$, so that the results of Section 12 would be applicable. As observed earlier, this would correspond to the so-called L^2 *regime* in the theory of GMC

In the single pass observation model, we again exploit our analysis of edge counts for this purpose. We recall that when $\nu = 0$, that is for the Poisson random geometric graph, $\mathcal{E}$ is a sum of indicators of all possible edges on the vertex set. Since edges are usually formed when the underlying points $x_i, x_j x_k$ are close to each other at the scale σ, and since $\sigma = O(n^{-1/d})$, we may conclude that $\mathcal{E}$ is a sum of a large (and Poisson) number of weakly dependent random variables. As such, it can be well-approximated by a compound Poisson random variable, which in turn admits a normal approximation with appropriate centering and scaling (c.f. (Penrose et al., 2003), (Van Der Hofstad, 2016)).

The upshot of this is that under $\nu = 0$, for large n, the normalized edge count $\frac{\mathcal{E}-\mathbb{E}[\mathcal{E}]}{\sqrt{\mathrm{Var}[\mathcal{E}]}}$ is approximately normally distributed. Under $\nu = 0$, the edge count is known to satisfy $\mathrm{Var}[\mathcal{E}] = C(d,\rho)n(1+o(1))$ (c.f. (Penrose et al., 2003)), which implies the approximate upper tail bound $\mathbb{P}[\mathcal{E} \geq C_2(0,d)\rho^2 n + t] \leq C\exp(-\frac{ct^2}{n})$. This suggests that, under $\nu = 0$, the probability $\mathbb{P}[\mathcal{E} \geq n^{1+\frac{1}{2}\nu_0}] \leq \exp(-cn^{1+\nu_0})$. On the other hand, under the alternative $\mathbb{E}[\mathcal{E}] = C(\gamma,d,\rho)n^{1+\nu}(1+o(1)) \geq C(\gamma,d,\rho)n^{1+\nu_0}(1+o(1)) \gg n^{1+\nu_0/2}$ as $n \to \infty$. This suggests that the threshold $n^{1+\nu_0/2}$ for the edge count separates the $\nu = 0$ and $\nu \geq \nu_0$ settings.

However, in our observation models, we do not have direct access to the latent size parameter n. Nonetheless, as discussed in Section 5, the observed network size N provides a good approximation of n upto an $O(1)$ multiplicative factor. Since $\mathcal{E}$ under the null and the alternative hypotheses are orders of magnitude (in n) apart (which is a consequence of the positive separation between the null and the alternative), we can use N as a substitute for n for obtaining a separation threshold.

Thus, in the single pass observation model,

$$\text{Declaring the presence of fractality if } \mathcal{E} > N^{1+\frac{1}{2}\nu_0} \quad (5)$$

would provide a detection procedure for fractality with good discriminatory power. In the multi pass observation model, we make use of the mean edge count $\overline{\mathcal{E}} = \frac{1}{m}\sum_{i=1}^m \mathcal{E}_i$ and the mean vertex count $\overline{N} = \frac{1}{m}\sum_{i=1}^m N_i$. In the regime of large m, they concentrate strongly around their respective means, with Gaussian CLT like effects. Thus, in the multi pass observation model

$$\text{Declaring the presence of fractality if } \overline{\mathcal{E}} > \overline{N}^{1+\frac{1}{2}\nu_0} \quad (6)$$

would provide a detection procedure for fractality with good discriminatory power.

6. Stochastic Block Models in the FGN paradigm

Stochastic Block Models (henceforth abbreviated as SBM) has become an important paradigm for understanding and investigation community structures in networks, social or otherwise. A long series of ground breaking results in this regard have been achieved in recent years; we refer the interested reader to ((Holland et al., 1983), (Abbe, 2017), (Abbe et al., 2015), (Abbe & Sandon, 2015), (Rácz et al., 2017), (Mukherjee, 2018)) for a partial overview of this vast and rapidly evolving field of research.

In the context of networks with fractal structures, it is natural to envisage a situation where there are multiple distinct communities in the network with potentially different fractal structures. It is also natural to posit that the communities have differing degrees of affinity to connect within each other as compared to connections across community boundaries, which might be rarer.

We encapsulate this idea in the form of a natural SBM structure in the context of the FGN model. We need the following ingredients:

- Two independent GMC-s $M^{\gamma_1}, M^{\gamma_2}$ corresponding to (possibly different) positive parameters γ_1, γ_2 respectively on the same domain $\Omega \subset \mathbb{R}^d$.
- Two different positive threshold parameters σ_{in} and σ_{out}.
- A size parameter $n \in \mathbb{N}$ and two independent Poisson random variables $N_1 \sim \mathrm{Poi}(nM^{\gamma_1}(\Omega))$ and $N_2 \sim \mathrm{Poi}(nM^{\gamma_2}(\Omega))$.

Given these ingredients, we construct the SBM on the FGN model as follows.

- We generate N_1 points $\{x_1, \ldots, x_{N_1}\}$ i.i.d. from the (normalized) measure M^{γ_1} and N_2 points $\{y_1, \ldots, y_{N_2}\}$ i.i.d. from the (normalized) measure M^{γ_2}.
- For each pair of points x_i, x_j, we connect them with an edge with probability $\propto \exp(-\frac{\|x_i - x_j\|^2}{\sigma_{\mathrm{in}}^2})$. Likewise, For each pair of points y_i, y_j, we connect them with an edge with probability $\propto \exp(-\frac{\|y_i - y_j\|^2}{\sigma_{\mathrm{in}}^2})$. These are the *intra-community links*.
- For each pair of points x_i, y_j, we connect them with an edge with probability $\propto \exp(-\frac{\|x_i - y_j\|^2}{\sigma_{\mathrm{out}}^2})$. These are the *inter-community links*.

We then forget the spatial identities of the points, and consider the resulting combinatorial graph $\mathcal{G}$, whose node set is the union of the node sets of the FGN-s $\mathcal{G}_1$ and $\mathcal{G}_2$, and whose edges are those of $\mathcal{G}_1 \cup \mathcal{G}_2$ along with the cross-community edges defined in the last step. This forms a natural SBM structure in the context of the FGN model.

A natural statistical question in this context would be to understand separation thresholds between in the intra-community connection radius σ_{in} and the inter-community connection radius σ_{out} which allow for detection of the different communities with reasonable accuracy and probabilistic guarantees, as the network size parameter $n \to \infty$. We leave this and related questions for future investigation.

7. Conclusion

We proposed and investigated a parametric statistical model of sparse random graphs called FGN that continuously interpolates between homogeneous, Poisson behavior on one hand, and fractal behavior with anomalous exponents and power law distributions on the other. We demonstrated how to construct a natural stochastic block model within the FGN framework. We investigated the fundamental questions of parameter estimation and detecting the presence of fractality based on observed network data.

This work raises many natural questions for further investigations. These include a more detailed and rigorous mathematical study of the FGN as a model of sparse random graphs. Another direction would be to obtain fundamental limits for natural statistical questions in this setting, particularly the Stochastic Block Model in this context, and investigating the computational-statistical trade-off for these problems. Extending our analytical results, and consequently the range of the estimation and detection procedures, beyond the L^2 regime of the GMC would be an natural and interesting question. From a modelling perspective, it would be natural to explore beyond Gaussianity in the construction of our networks, for which the basic motivation and the probabilistic fundamentals seem to be promising (see, e.g., (Barral & Mandelbrot, 2002), (Bacry & Muzy, 2003)). Another direction would be to venture beyond the Euclidean set-up as the latent space. We leave these and related questions as natural avenues for future investigation.

8. Acknowledgements

We are grateful to the referees for their careful reading of the manuscript, and for their illuminating comments and suggestions. SG was supported in part by the MOE grant R-146-000-250-133. KB was supported in part by UC Davis CeDAR (Center for Data Science and Artificial Intelligence Research) Innovative Data Science Seed Funding Program. XY was supported by the FNR Grant MISSILe (R-AGR-3410-12-Z) at Luxembourg and Singapore Universities.

References

Abbe, E. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

Abbe, E. and Sandon, C. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pp. 670–688. IEEE, 2015.

Abbe, E., Bandeira, A. S., and Hall, G. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2015.

Albert, R. and Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

Avnir, D., Biham, O., Lidar, D., and Malcai, O. Is the geometry of nature fractal? *Science*, 279(5347):39–40, 1998.

Bacry, E. and Muzy, J. F. Log-infinitely divisible multifractal processes. *Communications in Mathematical Physics*, 236(3):449–475, 2003.

Barabási, A.-L. et al. *Network science*. Cambridge university press, 2016.

Barral, J. and Mandelbrot, B. B. Multifractal products of cylindrical pulses. *Probability Theory and Related Fields*, 124(3):409–430, 2002.

Barrett, C. B. and Swallow, B. M. Fractal poverty traps. *World development*, 34(1):1–15, 2006.

Bassett, D. S., Meyer-Lindenberg, A., Achard, S., Duke, T., and Bullmore, E. Adaptive reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National Academy of Sciences*, 103(51): 19518–19523, 2006.

Batagelj, V. and Mrvar, A. A subquadratic triad census algorithm for large sparse networks with small maximum degree. *Social networks*, 23(3):237–243, 2001.

Benguigui, L. The fractal dimension of some railway networks. *Journal de Physique I*, 2(4):385–388, 1992.

Berestycki, N. Introduction to the gaussian free field and liouville quantum gravity. *Lecture notes*, 2015.

Berestycki, N. et al. An elementary approach to gaussian multiplicative chaos. *Electronic communications in Probability*, 22, 2017.

Bickel, P. J. and Chen, A. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50): 21068–21073, 2009.

Bickel, P. J. and Doksum, K. A. *Mathematical statistics: basic ideas and selected topics, volumes I-II package*. Chapman and Hall/CRC, 2015.

Bollobás, B., Kozma, R., and Miklos, D. *Handbook of large-scale random networks*, volume 18. Springer Science & Business Media, 2010.

Bubeck, S. and Ganguly, S. Entropic clt and phase transition in high-dimensional wishart matrices. *International Mathematics Research Notices*, 2018(2):588–606, 2018.

Bubeck, S., Mossel, E., and Rácz, M. Z. On the influence of the seed graph in the preferential attachment model. *IEEE Transactions on Network Science and Engineering*, 2(1):30–39, 2015.

Bubeck, S., Ding, J., Eldan, R., and Rácz, M. Z. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.

Bubeck, S., Eldan, R., Mossel, E., Rácz, M. Z., et al. From trees to seeds: on the inference of the seed from large trees in the uniform attachment model. *Bernoulli*, 23(4A): 2887–2916, 2017.

Caldarelli, G. *Scale-free networks: complex webs in nature and technology*. Oxford University Press, 2007.

Caldarelli, G., Marchetti, R., and Pietronero, L. The fractal properties of internet. *EPL (Europhysics Letters)*, 52(4): 386, 2000.

Caldarelli, G., Battiston, S., Garlaschelli, D., and Catanzaro, M. Emergence of complexity in financial networks. In *Complex Networks*, pp. 399–423. Springer, 2004.

Chung, F., Chung, F. R., Graham, F. C., Lu, L., Chung, K. F., et al. *Complex graphs and networks*. Number 107. American Mathematical Soc., 2006.

Claps, P., Fiorentino, M., and Oliveto, G. Informational entropy of fractal river networks. *Journal of Hydrology*, 187(1-2):145–156, 1996.

Crane, H. *Probabilistic foundations of statistical network analysis*. Chapman and Hall/CRC, 2018.

Daqing, L., Kosmidis, K., Bunde, A., and Havlin, S. Dimension of spatially embedded networks. *Nature Physics*, 7 (6):481–484, 2011.

De Florio, V., Bakhouya, M., Coronato, A., and Di Marzo, G. Models and concepts for socio-technical complex systems: towards fractal social organizations. *Systems Research and Behavioral Science*, 30(6):750–772, 2013.

de la Torre, S. R., Kalda, J., Kitt, R., and Engelbrecht, J. Fractal and multifractal analysis of complex networks: Estonian network of payments. *The European Physical Journal B*, 90(12):234, 2017.

Duchon, J., Robert, R., and Vargas, V. Forecasting volatility with the multifractal random walk model. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 22(1):83–108, 2012.

Duplantier, B. and Sheffield, S. Duality and the knizhnik-polyakov-zamolodchikov relation in liouville quantum gravity. *Physical Review Letters*, 102(15):150603, 2009.

Duplantier, B. and Sheffield, S. Liouville quantum gravity and kpz. *Inventiones mathematicae*, 185(2):333–393, 2011.

Duplantier, B., Rhodes, R., Sheffield, S., and Vargas, V. Log-correlated gaussian fields: an overview. *arXiv preprint arXiv:1407.5605*, 2014a.

Duplantier, B., Rhodes, R., Sheffield, S., and Vargas, V. Renormalization of critical gaussian multiplicative chaos and kpz relation. *Communications in Mathematical Physics*, 330(1):283–330, 2014b.

Duplantier, B., Rhodes, R., Sheffield, S., Vargas, V., et al. Critical gaussian multiplicative chaos: convergence of the derivative martingale. *The Annals of Probability*, 42 (5):1769–1808, 2014c.

Erdös, P., Rényi, A., et al. On random graphs. *Publicationes mathematicae*, 6(26):290–297, 1959.

Evertsz, C. J. Fractal geometry of financial time series. *Fractals*, 3(03):609–616, 1995.

Falconer, K. *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2004.

Franović, I. and Miljković, V. Percolation transition at growing spatiotemporal fractal patterns in models of mesoscopic neural networks. *Physical Review E*, 79(6):061923, 2009.

Fyodorov, Y. V., Le Doussal, P., and Rosso, A. Freezing transition in decaying burgers turbulence and random matrix dualities. *EPL (Europhysics Letters)*, 90(6):60004, 2010.

Gao, J., Hu, J., Mao, X., and Perc, M. Culturomics meets random fractal theory: insights into long-range correlations of social and natural phenomena over the past two centuries. *Journal of The Royal Society Interface*, 9(73): 1956–1964, 2012.

Gilbert, E. N. Random plane networks. *Journal of the society for industrial and applied mathematics*, 9(4):533–543, 1961.

Goh, K.-I., Salvi, G., Kahng, B., and Kim, D. Skeleton and fractal scaling in complex networks. *Physical review letters*, 96(1):018701, 2006.

Grover, A. and Leskovec, J. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864, 2016.

Guédon, O. and Vershynin, R. Community detection in sparse networks via grothendieck's inequality. *Probability Theory and Related Fields*, 165(3-4):1025–1049, 2016.

Haight, F. A. Handbook of the poisson distribution. 1967.

Hill, R. A., Bentley, R. A., and Dunbar, R. I. Network scaling reveals consistent fractal pattern in hierarchical mammalian societies. *Biology letters*, 4(6):748–751, 2008.

Høegh-Krohn, R. A general class of quantum fields without cut-offs in two space-time dimensions. *Communications in Mathematical Physics*, 21(3):244–255, 1971.

Holland, P. W., Laskey, K. B., and Leinhardt, S. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

Inaoka, H., Ninomiya, T., Taniguchi, K., Shimizu, T., and Takayasu, H. Fractal network derived from banking transaction– an analysis of network structures formed by financial institutions, 2004.

Jackson, M. O. *Social and economic networks*. Princeton university press, 2010.

Johnson, D. B. Efficient algorithms for shortest paths in sparse networks. *Journal of the ACM (JACM)*, 24(1): 1–13, 1977.

Kahane, J.-P. Sur le chaos multiplicatif. *Ann. Sci. Math. Québec*, 9(2):105–150, 1985.

Kahane, J.-P. and Peyriere, J. Sur certaines martingales de benoit mandelbrot. *Advances in mathematics*, 22(2): 131–145, 1976.

Kim, J. S., Goh, K.-I., Kahng, B., and Kim, D. Fractality and self-similarity in scale-free networks. *New Journal of Physics*, 9(6):177, 2007.

Kolmogorov, A. N. The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. *Cr Acad. Sci. URSS*, 30:301–305, 1941.

Kolmogorov, A. N. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high reynolds number. *Journal of Fluid Mechanics*, 13(1):82–85, 1962.

Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L., and Zhang, P. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

La Barbera, P. and Rosso, R. On the fractal dimension of stream networks. *Water Resources Research*, 25(4): 735–741, 1989.

Lacoin, H. A short course on real and complex gaussian multiplicative chaos. 2019.

Lee, K., Hong, S., Kim, S. J., Rhee, I., and Chong, S. Slaw: A new mobility model for human walks. In *IEEE INFOCOM 2009*, pp. 855–863. IEEE, 2009.

Lee, K., Hong, S., Kim, S. J., Rhee, I., and Chong, S. Slaw: self-similar least-action human walk. *IEEE/ACM Transactions On Networking*, 20(2):515–529, 2011.

Lewis, T. G. *Network science: Theory and applications*. John Wiley & Sons, 2011.

Liu, Y., Gopikrishnan, P., Stanley, H. E., et al. Statistical properties of the volatility of price fluctuations. *Physical review e*, 60(2):1390, 1999.

Lovász, L. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

Lyudmyla, K., Vitalii, B., and Tamara, R. Fractal time series analysis of social network activities. In *2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T)*, pp. 456–459. IEEE, 2017.

Mandelbrot, B. B. *The fractal geometry of nature*, volume 173. WH freeman New York, 1983.

Mandelbrot, B. B. *Fractals and scaling in finance: Discontinuity, concentration, risk. Selecta volume E*. Springer Science & Business Media, 2013.

Mandelbrot, B. B. and Hudson, R. L. *The (mis) behaviour of markets: a fractal view of risk, ruin and reward*. Profile books, 2010.

Mezard, M., Mezard, M., and Montanari, A. *Information, physics, and computation*. Oxford University Press, 2009.

Mossel, E., Neeman, J., and Sly, A. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.

Mukherjee, S. S. *On some inference problems for networks*. PhD thesis, UC Berkeley, 2018.

Murcio, R., Masucci, A. P., Arcaute, E., and Batty, M. Multifractal to monofractal evolution of the london street network. *Physical Review E*, 92(6):062130, 2015.

Orbach, R. Dynamics of fractal networks. *Science*, 231 (4740):814–819, 1986.

Orbanz, P. and Roy, D. M. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE transactions on pattern analysis and machine intelligence*, 37(2):437–461, 2014.

Palla, G., Lovász, L., and Vicsek, T. Multifractal network generator. *Proceedings of the National Academy of Sciences*, 107(17):7640–7645, 2010.

Pavón-Domínguez, P., Ariza-Villaverde, A. B., Rincón-Casado, A., de Ravé, E. G., and Jiménez-Hornero, F. J. Fractal and multifractal characterization of the scaling geometry of an urban bus-transport network. *Computers, Environment and Urban Systems*, 64:229–238, 2017.

Penrose, M. et al. *Random geometric graphs*, volume 5. Oxford university press, 2003.

Pook, W. and Janßen, M. Multifractality and scaling in disordered mesoscopic systems. *Zeitschrift für Physik B Condensed Matter*, 82(2):295–298, 1991.

Rácz, M. Z. and Richey, J. A smooth transition from wishart to goe. *Journal of Theoretical Probability*, 32(2):898–906, 2019.

Rácz, M. Z., Bubeck, S., et al. Basic models and questions in statistical network analysis. *Statistics Surveys*, 11:1–47, 2017.

Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S. J., and Chong, S. On the levy-walk nature of human mobility. *IEEE/ACM transactions on networking*, 19(3):630–643, 2011.

Rhodes, R., Vargas, V., et al. Gaussian multiplicative chaos and applications: a review. *Probability Surveys*, 11, 2014.

Rhodes, R. et al. Lecture notes on gaussian multiplicative chaos and liouville quantum gravity. *arXiv preprint arXiv:1602.07323*, 2016.

Rinaldo, A., Rodriguez-Iturbe, I., Rigon, R., Bras, R. L., Ijjasz-Vasquez, E., and Marani, A. Minimum energy and fractal structures of drainage networks. *Water Resources Research*, 28(9):2183–2195, 1992.

Rinaldo, A., Rodriguez-Iturbe, I., Rigon, R., Ijjasz-Vasquez, E., and Bras, R. Self-organized fractal river networks. *Physical review letters*, 70(6):822, 1993.

Salingaros, N. A. Connecting the fractal city. *5th Biennial of towns and town planners in Europe, Barcelona*, 2003.

Sarkar, P. and Moore, A. W. Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems*, pp. 1145–1152, 2006.

Sheffield, S. Gaussian free fields for mathematicians. *Probability theory and related fields*, 139(3-4):521–541, 2007.

Simon, B. $P(\Phi)_2$ *Euclidean (Quantum) Field Theory*. Princeton University Press, 2015.

Song, C., Havlin, S., and Makse, H. A. Self-similarity of complex networks. *Nature*, 433(7024):392–395, 2005.

Song, C., Havlin, S., and Makse, H. A. Origins of fractality in the growth of complex networks. *Nature physics*, 2(4): 275–281, 2006.

Spielman, D. A. Graphs and networks. *Lecture Notes*, 3, 2010.

Strogatz, S. H. Exploring complex networks. *Nature*, 410 (6825):268–276, 2001.

Van Der Hofstad, R. *Random graphs and complex networks*, volume 1. Cambridge university press, 2016.

Watts, D. J. *Six degrees: The science of a connected age*. WW Norton & Company, 2004.

Wigner, E. P. The unreasonable effectiveness of mathematics in the natural sciences. In *Mathematics and Science*, pp. 291–306. World Scientific, 1990.

9. Additional Simulations

In this section, we provide surface-plots of the GMC measure as the parameter ν varies.

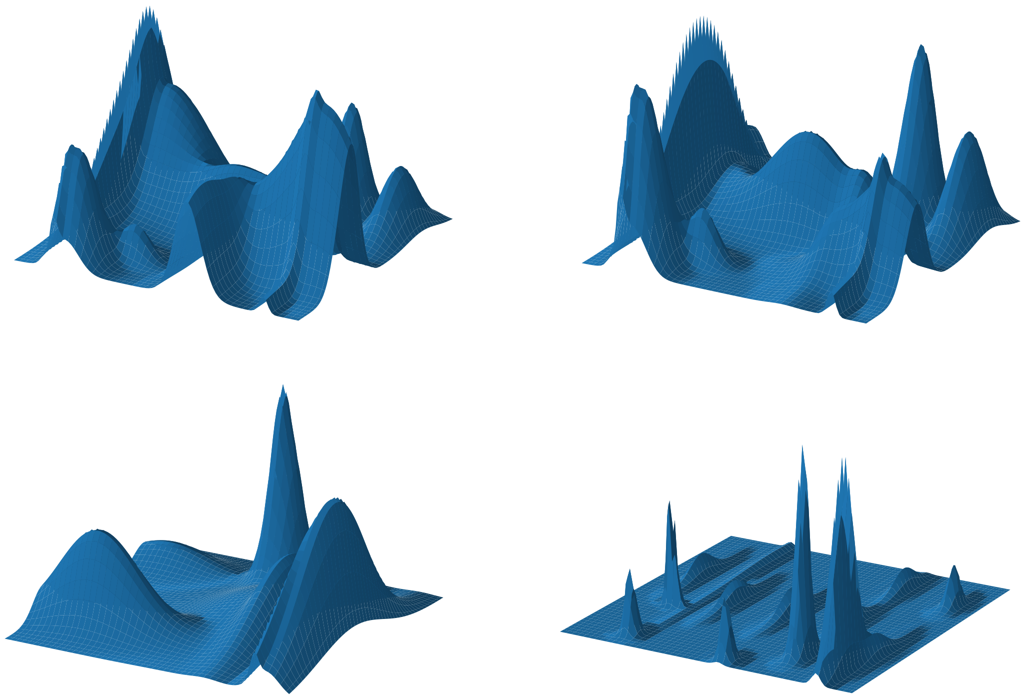

Figure 3. As we move from top to bottom and from left to right, the value of ν increases.

10. Gaussian Multiplicative Chaos : Fundamentals

In this section, we provide a more technical introduction to GMC, introducing tools which will aid in our analytical investigations subsequently. For an elaborate discussion, we refer the reader to the extensive accounts (Rhodes et al., 2014), (Rhodes et al., 2016), (Berestycki, 2015), (Berestycki et al., 2017), (Lacoin, 2019) for a partial list, and the references contained therein. The introduction of a Brownian motion below is helpful for computations, for more detail we refer the interested reader to (Duplantier et al., 2014c), (Duplantier et al., 2014b) and (Duplantier et al., 2014a).

Let $\{X_t(x), t \geq 0, x \in \mathbb{R}^d\}$ be a centered Gaussian field, which is a standard Brownian motion for each fixed x and

$$\mathbb{E}[X_t(x)X_t(y)] = \int_1^{e^t} \frac{k(u(x-y))}{u} du, \tag{7}$$

therefore stationary in space variable. We make the following assumptions

- The map $k : \mathbb{R}^d \to [0, \infty)$ is radial, i.e. $k(x) = k(|x|e)$ for any x and unit vector $e \in \mathbb{R}^d$.
- $k(0) = 1$
- $k \in C^1$ and decays fast enough at infinity such that $\int_1^{\infty} \frac{k(u)}{u} du < \infty$.

As $t \to \infty$, one obtains a log-correlated Gaussian field X as a random distribution. It turns out that such functions k lead to the limiting covariance function of the Gaussian field X have the following form :

$$K(x, y) = \ln_+ \frac{T}{|x-y|} + g(x-y),$$

where $T > 0$ and g is a bounded continuous function. With these ingredients in hand, we may define

$$M^\gamma := \lim_{t\to\infty} M_t^\gamma \text{ a.s., where } M_t^\gamma(\mathrm{d}x) = e^{\gamma X_t(x) - \frac{\gamma^2}{2}\mathbb{E}[X_t(x)^2]}\mathrm{d}x, \tag{8}$$

and the convergence is guaranteed by a martingale structure that is known to be inherent in this setting. Since, for each x, the process $X_t(x)$ is a Brownian motion, we have $\mathbb{E}[X_t(x)^2] = t$, and we may write

$$M_t^\gamma(\mathrm{d}x) = e^{\gamma X_t(x) - \frac{\gamma^2 t}{2}}\mathrm{d}x.$$

If $\gamma^2 < 2d$ (equivalently, $\nu = \frac{\gamma^2}{d} < 2$), the limit M^γ is a non-degenerate measure, otherwise M^γ is a trivial zero measure. This regime $\nu < 2$ where the GMC is a non-degenerate measure will be referred to as the subcritical regime. In our analytical considerations, we will assume the GMC is subcritical and we consider the GMC on the d-dimensional unit cube $\Omega = [-1/2, 1/2]^d$.

11. Determining the connectivity threshold σ and *sparse* random graphs

In this section, we determine the *right regime* of the connectivity threshold σ. In doing so, our guiding principle would be to obtain a sparse random graph model in the end, one in which the number of neighbours from the FGN of a given point in the latent social space is typically $O(1)$.

To this end, consider the FGN with $N \sim \mathrm{Poi}(nM^\gamma(\Omega))$, nodes $\{x_1, \dots, x_N\}$ and threshold σ. Consider a point $x_0 \in \Omega$. By the notation $x \sim y$, we mean that the point x is connected to the point y by an edge. Observe that, under our connection model for edge formation (once we are given some nodes), $\mathbb{P}[x_0 \sim x_i] = e^{-\frac{|x_i - x_0|^2}{\sigma^2}}$, and the total number of points to which x_0 may be connected to in this manner is $\left(\sum_{i=1}^N \mathbb{1}_{x_0 \sim x_i}\right)$. Therefore, in the regime of small connection threshold σ, since

$$\int_{\mathbb{R}^d} e^{-|x|^2}\mathrm{d}x = \pi^{\frac{d}{2}},$$

the expected number of points in this FGN that would be connected of x_0 is :

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^N \mathbb{1}_{x_0\sim x_i}\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^N \mathbb{1}_{x_0\sim x_i} \big| \text{FGN}\right]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^N e^{-\frac{|x_i-x_0|^2}{\sigma^2}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^N e^{-\frac{|x_i-x_0|^2}{\sigma^2}} \big| N, \mathrm{d}M^\gamma\right]\right] \\
&= \mathbb{E}\left[N \cdot \int_\Omega e^{-\frac{|x-x_0|^2}{\sigma^2}} \frac{M^\gamma(\mathrm{d}x)}{M^\gamma(\Omega)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[N \cdot \int_\Omega e^{-\frac{|x-x_0|^2}{\sigma^2}} \frac{M^\gamma(\mathrm{d}x)}{M^\gamma(\Omega)} \big| \mathrm{d}M^\gamma\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[N \big| \mathrm{d}M^\gamma\right] \cdot \int_\Omega e^{-\frac{|x-x_0|^2}{\sigma^2}} \frac{M^\gamma(\mathrm{d}x)}{M^\gamma(\Omega)}\right] \\
&= n \cdot \mathbb{E}\left[\int_\Omega e^{-\frac{|x-x_0|^2}{\sigma^2}} M^\gamma(\mathrm{d}x)\right] \\
&= n \cdot \left[\int_\Omega e^{-\frac{|x-x_0|^2}{\sigma^2}} \mathrm{d}x\right] \\
&= n\sigma^d \cdot \int_{\Omega/\sigma} e^{-\frac{|x-x_0|^2}{\sigma^2}} \mathrm{d}x
\end{aligned}$$

$$= n(\sqrt{\pi}\sigma)^d(1 + o(1)),$$

where, the fourth equality follows since the $\{x_i\}_{i=1}^N$ are i.i.d. $\mathrm{d}M^\gamma$ given $N, \mathrm{d}M^\gamma$, the seventh equality follows since $\mathbb{E}[N|\mathrm{d}M^\gamma] = nM^\gamma(\Omega)$, and, the eighth inequality follows since the expected measure $\mathbb{E}[\mathrm{d}M^\gamma(x)]$ is Lebesgue. We summarize this as follows.

Theorem 11.1. *In the FGN model with size parameter n, setting the threshold parameter*

$$\sigma = \frac{1}{\sqrt{\pi}}\rho^{1/d}n^{-1/d},$$

one has that the expected number of neighbours of a given point is asymptotically $\rho \in (0, \infty)$.

We call ρ the *density parameter* or the *degree parameter* of the FGN.

12. The statistics of small subgraph counts in the FGN : an analysis of edge counts

In this section, we investigate the statistic of edge counts in the FGN model. To this end, consider the FGN with $N \sim \mathrm{Poi}(nM^\gamma(\Omega))$, nodes $\{x_1, \ldots, x_N\}$ and threshold σ. Let $\mathcal{E}$ denote the number of edges in this FGN.

In this section, we will work in the setting $\gamma^2 < d$ (equivalently, $\nu < 1$) for our analysis. Interestingly, this corresponds to the so-called L^2 regime of the GMC, where the model is believed to be technically more tractable in relative terms. We believe that similar results would be true for the full range of validity of the GMC and the FGN model (i.e., all the way up to $\nu < 2$), albeit technically more challenging. We leave this as an interesting direction for future study.

In the computations that follow, we will use the fact that if $\Lambda \sim \mathrm{Poi}(\lambda)$, then the second *factorial moment* of Λ is given by the relation $\mathbb{E}[\binom{\Lambda}{2}] = \frac{\lambda^2}{2!}$ (c.f. (Haight, 1967)). Consequently, recalling that given the GMC $\mathrm{d}M^\gamma$ the node count $N \sim \mathrm{Poi}(nM^\gamma(\Omega))$, we may deduce that $\mathbb{E}[\binom{N}{2}] = \frac{1}{2} \cdot n^2 M^\gamma(\Omega)^2$.

In view of this, we may proceed as

$$\begin{aligned}
\mathbb{E}[\mathcal{E}] =& \mathbb{E}\left[\mathbb{E}[\mathcal{E}|\text{FGN}]\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[\sum_{1\le i<j\le N} \mathbb{1}_{x_i\sim x_j} \big| \text{FGN}\right]\right] \\
=& \mathbb{E}\left[\sum_{1\le i<j\le N} e^{-|x_i-x_j|^2/\sigma^2}\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[\sum_{1\le i<j\le N} e^{-|x_i-x_j|^2/\sigma^2} \big| N, \mathrm{d}M^\gamma\right]\right] \\
=& \mathbb{E}\left[\binom{N}{2} \cdot \iint_{\Omega\times\Omega} e^{-|x-y|^2/\sigma^2} \frac{M^\gamma(\mathrm{d}x)M^\gamma(\mathrm{d}y)}{M^\gamma(\Omega)^2}\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[\binom{N}{2} \cdot \iint_{\Omega\times\Omega} e^{-|x-y|^2/\sigma^2} \frac{M^\gamma(\mathrm{d}x)M^\gamma(\mathrm{d}y)}{M^\gamma(\Omega)^2} \big| \mathrm{d}M^\gamma\right]\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[\binom{N}{2} \big| \mathrm{d}M^\gamma\right] \cdot \iint_{\Omega\times\Omega} e^{-|x-y|^2/\sigma^2} \frac{M^\gamma(\mathrm{d}x)M^\gamma(\mathrm{d}y)}{M^\gamma(\Omega)^2}\right] \\
=& \mathbb{E}\left[\frac{1}{2}n^2 M^\gamma(\Omega)^2 \cdot \iint_{\Omega\times\Omega} e^{-|x-y|^2/\sigma^2} \frac{M^\gamma(\mathrm{d}x)M^\gamma(\mathrm{d}y)}{M^\gamma(\Omega)^2}\right] \\
=& \frac{n^2}{2}\mathbb{E}\left[\iint_{\Omega\times\Omega} e^{-|x-y|^2/\sigma^2} M^\gamma(\mathrm{d}x)M^\gamma(\mathrm{d}y)\right].
\end{aligned}$$

where the fifth equality follows since the $\{x_i\}_{i=1}^N$ are i.i.d. $\mathrm{d}M^\gamma$ given $N, \mathrm{d}M^\gamma$. For further analysis, we consider

$$I := \mathbb{E}\left[\iint_{\Omega^2} \exp\left(-\frac{|x-y|^2}{\sigma^2}\right) M^\gamma(\mathrm{d}x)M^\gamma(\mathrm{d}y)\right].$$

and

$$I_t := \mathbb{E}\left[\iint_{\Omega^2} \exp\left\{-\frac{|x-y|^2}{\sigma^2} + \gamma(X_t(x)+X_t(y)+X_t(z)) - \gamma^2 t\right\} \mathrm{d}x\mathrm{d}y\right].$$

In view of the convergence (8), we will use I_t as an approximation for I as $t \to \infty$.

Using the fact that for fixed t the field $\{X_t(x)\}$ is a centered Gaussian random field with covariance structure(7), we may then proceed further as

$$\begin{aligned}
\mathbb{E}[\mathcal{E}] =& \frac{n^2}{2} \lim_{t\to\infty} \mathbb{E}\left[\iint_{\Omega\times\Omega} \exp\left\{-\frac{|x-y|^2}{\sigma^2} + \gamma(X_t(x)+X_t(y)) - \gamma^2 t\right\} \mathrm{d}x\mathrm{d}y\right] \\
=& \frac{n^2}{2} \lim_{t\to\infty} \iint_{\Omega\times\Omega} \exp\left\{-\frac{|x-y|^2}{\sigma^2} - \gamma^2 t\right\} \mathbb{E}\left[\exp\left(\gamma(X_t(x)+X_t(y))\right)\right] \mathrm{d}x\mathrm{d}y \\
=& \frac{n^2}{2} \lim_{t\to\infty} \iint_{\Omega\times\Omega} \exp\left\{-\frac{|x-y|^2}{\sigma^2} - \gamma^2 t\right\} \cdot \exp\left\{\frac{\gamma^2}{2}(2t + 2\int_1^{e^t} \frac{k(u(x-y))}{u}\mathrm{d}u)\right\} \mathrm{d}x\mathrm{d}y \\
=& \frac{n^2}{2} \lim_{t\to\infty} \iint_{\Omega\times\Omega} \exp\left\{-\frac{|x-y|^2}{\sigma^2} + \gamma^2 \int_1^{e^t} \frac{k(u(x-y))}{u}\mathrm{d}u\right\} \mathrm{d}x\mathrm{d}y \\
=& \frac{n^2}{2} \iint_{\Omega\times\Omega} \exp\left\{-\frac{|x-y|^2}{\sigma^2} + \gamma^2 \int_1^{\infty} \frac{k(u(x-y))}{u}\mathrm{d}u\right\} \mathrm{d}x\mathrm{d}y. \qquad (9)
\end{aligned}$$

A change of variables shows that

$$\int_1^\infty \frac{k(ux)}{u}\mathrm{d}u = \int_{|x|}^\infty \frac{k(u)}{u}\mathrm{d}u =: \phi(|x|). \qquad (10)$$

Combining (9) and (10), together with another change of variables $(x,y) \mapsto (x/\sigma, y/\sigma)$, gives

$$\mathbb{E}[\mathcal{E}] = \frac{n^2\sigma^{2d}}{2} \iint_{\Omega/\sigma\times\Omega/\sigma} \exp\Big\{-|x-y|^2 + \gamma^2\phi\left(|x-y|\sigma\right)\Big\}\mathrm{d}x\mathrm{d}y \qquad (11)$$

where $\Omega/\sigma = [-1/2\sigma, 1/2\sigma]^d$. Since $|x-y|\sigma \le 1$, one has

$$\phi(|x-y|\sigma) = \int_{|x-y|\sigma}^1 \frac{k(u)}{u}\mathrm{d}u + \int_1^\infty \frac{k(u)}{u}\mathrm{d}u. \qquad (12)$$

where the second term is finite by assumption.

Thus, as $\sigma \downarrow 0$, one obtains from (12)

$$\phi(|x-y|\sigma) = \left(\int_{|x-y|\sigma}^1 \frac{k(u)}{u}\mathrm{d}u\right)(1+o(1)) = \left(\log\frac{1}{|x-y|} + \log\frac{1}{\sigma}\right)(1+o(1)). \qquad (13)$$

Combining (11) with (13), in the regime of small σ and $\gamma^2 < d$ (so that $\left(\int_{\mathbb{R}^d} \frac{1}{|u|^{\gamma^2}} e^{-|u|^2}\mathrm{d}u\right) < \infty$), we obtain

$$\begin{aligned}
\mathbb{E}[\mathcal{E}] =& \frac{n^2\sigma^{2d}}{2} \cdot \iint_{\Omega/\sigma\times\Omega/\sigma} \exp\left(-|x-y|^2\right) \cdot \frac{1}{|x-y|^{\gamma^2}\sigma^{\gamma^2}} \mathrm{d}x\mathrm{d}y \\
=& \frac{n^2}{2}\sigma^{2d-\gamma^2} \cdot \int_{\Omega/\sigma} \left(\int_{\substack{u=y-x \\ y\in\Omega/\sigma}} \frac{1}{|u|^{\gamma^2}} e^{-|u|^2}\mathrm{d}u\right)\mathrm{d}x \\
=& \frac{n^2}{2}\sigma^{2d-\gamma^2} \cdot \left(\int_{\Omega/\sigma} \left(\int_{\mathbb{R}^d} \frac{1}{|u|^{\gamma^2}} e^{-|u|^2}\mathrm{d}u\right)\mathrm{d}x\right) \cdot (1+o(1)) \\
=& \frac{n^2}{2}\sigma^{2d-\gamma^2} \cdot \left(\int_{\mathbb{R}^d} \frac{1}{|u|^{\gamma^2}} e^{-|u|^2}\mathrm{d}u\right) \cdot \sigma^{-d} \cdot (1+o(1))
\end{aligned}$$

$$=C_1(\gamma, d) \cdot n^2 \sigma^{d-\gamma^2} \left(1 + o(1)\right), \tag{14}$$

where,

$$C_1(\gamma, d) = \frac{1}{2} \int_{\mathbb{R}^d} \frac{1}{|u|^{\gamma^2}} e^{-|u|^2} \mathrm{d}u < \infty.$$

Using the choice $\sigma = \frac{1}{\sqrt{\pi}} \rho^{1/d} n^{-1/d}$ and $\nu = \gamma^2/d$, we finally obtain

$$\mathbb{E}[\mathcal{E}] = C(\gamma, d) \rho^{1-\nu} n^{1+\nu} \left(1 + o(1)\right). \tag{15}$$

We record our analysis as follows.

Theorem 12.1. *In the FGN model with size parameter n, density parameter ρ and fractality parameter $\nu = \gamma^2/d < 1$, the expected edge count satisfies*

$$\mathbb{E}[\mathcal{E}] = C(\gamma, d) \rho^{1-\nu} n^{1+\nu} \left(1 + o(1)\right)$$

as $n \to \infty$, where $C(\gamma, d) = \frac{1}{2} \pi^{\frac{1}{2} d - \frac{1}{2} \gamma^2} \cdot \left(\int_{\mathbb{R}^d} \frac{1}{|u|^{\gamma^2}} e^{-|u|^2} \mathrm{d}u \right)$.