ARTICLE IN PRESS
YINCO:4359

Information and Computation ••• (••••) •••–•••

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Symmetric graph properties have independent edges

Dimitris Achlioptas [a,b,1,2], Paris Siminelakis [c,∗,1,3]

[a] Department of Computer Science, University of California, Santa Cruz, United States
[b] Department of Informatics & Telecommunications, University of Athens, Greece
[c] Department of Electrical Engineering, Stanford University, United States

ARTICLE INFO

Article history:
Received 6 July 2015
Available online xxxx

Keywords:
Random graphs
Symmetric graph properties
Concentration of measure
Approximate independence
Maximum Entropy Principle

ABSTRACT

In the study of random structures we often face a trade-off between realism and tractability, the latter typically enabled by independence assumptions. In this work we initiate an effort to bridge this gap by developing tools that allow us to work with independence without assuming it. Let $\mathcal{G}_n$ be the set of all graphs on n vertices and let S be an arbitrary subset of $\mathcal{G}_n$, e.g., the set of all graphs with m edges. The study of random networks can be seen as the study of properties that are true for *most* elements of S, i.e., that are true with high probability for a uniformly random element of S. With this in mind, we pursue the following question: *What are general sufficient conditions for the uniform measure on a set of graphs $S \subseteq \mathcal{G}_n$ to be well-approximable by a product measure on the set of all possible edges?*

1. Introduction

Since their introduction in 1959 by Erdős and Rényi [1] and Gilbert [2], respectively, $G(n,m)$ and $G(n,p)$ random graphs have dominated the mathematical study of random networks [3,4]. Given n vertices, $G(n,m)$ selects uniformly among all graphs with m edges, whereas $G(n,p)$ includes each edge independently with probability p. A refinement of $G(n,m)$ are graphs chosen uniformly among all graphs with a given degree sequence, a distribution made tractable by the configuration model of Bollobás [3]. Due to their mathematical tractability these three models have become a cornerstone of Probabilistic Combinatorics and have found application in the Analysis of Algorithms, Coding Theory, Economics, Game Theory, and Statistical Physics.

At the foundation of this mathematical tractability lies symmetry: the probability of all edge-sets of a given size is either the same, as in $G(n,p)$ and $G(n,m)$, or merely a function of the potency of the vertices involved, as in the configuration model. This extreme symmetry bestows numerous otherworldly properties, including near-optimal expansion. Perhaps most importantly, it amounts to a complete lack of geometry, as manifest by the fact that the shortest path metric of such graphs suffers maximal distortion when embedded in Euclidean space [5]. In contrast, vertices of real networks are typically em-

∗ Corresponding author.
E-mail addresses: optas@cs.ucsc.edu (D. Achlioptas), psimin@stanford.edu (P. Siminelakis).

[1] Research supported by ERC Starting Grant StG-210743.
[2] Research supported by NSF Grant CCF-1514128, an Alfred P. Sloan Fellowship, the European Union (European Social Fund ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) – Research Funding Program: ARISTEIA II.
[3] Supported by an Onassis Foundation Scholarship.

https://doi.org/10.1016/j.ic.2018.02.017
0890-5401/

Please cite this article in press as: D. Achlioptas, P. Siminelakis, Symmetric graph properties have independent edges, Inf. Comput. (2018), https://doi.org/10.1016/j.ic.2018.02.017

bedded in some low-dimensional geometry, either explicit (physical networks), or implicit (social and other latent semantics networks), with distance being a strong factor in determining the probability of edge formation.

While the shortcomings of the classical models have long been recognized, proposing more realistic models is not an easy task. The difficulty lies in achieving a balance between realism and mathematical tractability: it is only too easy to create network models that are both ad hoc and intractable. By now there are thousands of papers proposing different ways to generate graphs with desirable properties [6] the vast majority of which only provide heuristic arguments to support their claims. For a gentle introduction the reader is referred to the book of Newman [7] and for a more mathematical treatment to the books of Chung and Lu [8] and of Durrett [9].

In trying to replicate real networks one approach is to keep adding features, creating increasingly complicated models, in the hope of *matching* observed properties. Ultimately, though, the purpose of any good model is prediction. In that sense, the reason to study (random) graphs with certain properties is to understand what *other* graph properties are (typically) implied by the assumed properties. For instance, the reason we study the uniform measure on graphs with m edges, i.e., $G(n,m)$, is to understand "what properties are typically implied by the property of having m edges" (and we cast the answer as "properties that hold with high probability in a 'random' graph with m edges"). Notably, analyzing the uniform measure even for this simplest property is non-trivial. The reason is that it entails the single massive choice of an m-subset of edges, rather than m independent choices. In contrast, the independence of choices in $G(n,p)$ makes that distribution far more accessible, dramatically enabling analysis.

Connecting $G(n,m)$ and $G(n,p)$ is a classic result of random graph theory. The key observation is that to sample according to $G(n,p)$, since edges are independent and equally likely, we can first sample an integer $m \sim \mathrm{Bin}\left(\binom{n}{2}, p\right)$ and then sample a uniformly random graph with m edges, i.e., $G(n,m)$. Thus, for $p = p(m) = m/\binom{n}{2}$, the random graph $G \sim G(n,m)$ and the two random graphs $G^{\pm} \sim G(n,(1\pm\epsilon)p)$ can be coupled so that, viewing each graph as a set of edges, with high probability,

$$G^- \subseteq G \subseteq G^+ \ . \tag{1}$$

The significance of this relationship between what we *wish* to study (uniform measure) and what we *can* study (product measure) can not be overestimated. It manifests most dramatically in the study of monotone properties: to study a monotone, say, increasing property in $G \sim G(n,m)$ it suffices to bound from above its probability in G^+ and from below in G^-. This connection has been thoroughly exploited to establish threshold functions for a host of monotone graph properties such as Connectivity, Hamiltonicity, and Subgraph Existence, making it the workhorse of random graph theory.

In this work we seek to extend the above relationship between the uniform measure and product measures to properties more delicate than having a given number of edges. In doing so we (i) provide a tool that can be used to revisit a number of questions in random graph theory from a more realistic angle and (ii) lay the foundation for designing random graph models eschewing independence assumptions. For example, our tool makes short work of the following set of questions (which germinated our work):

Given an arbitrary collection of n points on the plane what can be said about the set of all graphs that can be built on them using a given amount of wire, i.e., when connecting two points consumes wire equal to their distance? What does a uniformly random such graph look like? How does it change as a function of the available wire?

1.1. Our contribution

A product measure on the set of all undirected simple graphs on n vertices, $\mathcal{G}_n$, is specified by a symmetric matrix $\mathbf{Q} \in [0,1]^{n\times n}$ where $Q_{ii} = 0$ for $i \in [n]$. By analogy to $G(n,p)$ we denote by $G(n,\mathbf{Q})$ the measure in which every edge $\{i,j\}$ is included independently with probability $Q_{ij} = Q_{ji}$. Let $S \subseteq \mathcal{G}_n$ be arbitrary. Our main result is a sufficient condition for the uniform measure on S, denoted by $U(S)$, to be approximable by a product measure in the following sense.

Sandwichability. *The measure $U(S)$ is (ϵ,δ)-sandwichable if there exists an $n\times n$ symmetric matrix $\mathbf{Q}$ such that the distributions $G \sim U(S)$ and $G^{\pm} \sim G(n,(1\pm\epsilon)\mathbf{Q})$ can be coupled so that $\mathbb{P}[G^- \subseteq G \subseteq G^+] \geq 1-\delta$.*

Informally, the two conditions required for our theorem to hold are:

Partition symmetry. The set S should be symmetric with respect to *some* partition $\mathcal{P} = (P_1,\ldots,P_k)$ of the $\binom{n}{2}$ possible edges. More specifically, for a partition $\mathcal{P}$ define the *edge profile* of a graph G with respect to $\mathcal{P}$ to be the k-dimensional vector $\mathbf{m}(G) := (m_1(G),\ldots,m_k(G))$ where $m_i(G)$ counts the number of edges in G from part P_i. Partition symmetry amounts to the requirement that the characteristic function of S can depend on *how many* edges are included from each part but not on *which* edges. That is, if we let $\mathbf{m}(S) := \{\mathbf{m}(G) : G \in S\}$, then $\forall G \in \mathcal{G}_n, \mathbb{I}_S(G) = \mathbb{I}_{\mathbf{m}(S)}(\mathbf{m}(G))$. The $G(n,m)$ model is recovered by considering the trivial partition with $k=1$ parts and $\mathbf{m}(S) = \{m\}$. Far more interestingly, in our motivating example edges are partitioned into equivalence classes according to their cost $\mathbf{c}$ (distance of endpoints) and the characteristic function allows graphs whose edge profile $\mathbf{m}(G)$ does not violate the total wire budget $C_B = \{\mathbf{v} \in \mathbb{N}^k : \mathbf{c}^\intercal \mathbf{v} \leq B\}$. We discuss the motivation for edge-partition symmetry at length in Section 2.