Trade-Offs between Density and Robustness in Random Interconnection Graphs[*]

Philippe Flajolet[1], Kostas Hatzis[2,3], Sotiris Nikoletseas[2,3], and Paul Spirakis[2,3]

[1] Algorithms Project, INRIA Rocquencourt, France
Philippe.Flajolet@inria.fr
[2] Computer Technology Institute (CTI), Patras, Greece
{hatzis, nikole, spirakis}@cti.gr
[3] Computer Engineering and Informatics Department,
Patras University, Greece.

Abstract. Graphs are models of communication networks. This paper applies combinatorial and symbolic-analytic techniques in order to characterize the interplay between two parameters of a random graph: its density (the number of edges in the graph) and its robustness to link failures, where robustness here means multiple connectivity by short disjoint paths. A triple (G, s, t), where G is a graph and s, t are designated vertices, is called $\ell - robust$ if s and t are connected via at least two edge-disjoint paths of length at most ℓ. We determine here the expected number of ways to get from s to t via two edge-disjoint paths in the random graph model $G_{n,p}$. We then derive bounds on related threshold probabilities $p_{n,\ell}$ as functions of ℓ and n.

1 Introduction

In recent years the development and use of communication networks has increased dramatically. In such networks, basic physical architecture combined with traffic congestion or operating system decisions, result in a certain, dynamically changing geometry of the graph of interconnections. We adopt the random graph model of $G_{n,p}$ (see [3], [4]) to capture link availability with probability p (independently for each link) in a fully connected graph. Even in such a simple network model, it is interesting to investigate the trade-off between its *density* (the number of edges) and its *robustness* to link failures. The existence of alternative paths in such graphs models desired reliability and efficiency properties, such as the ability to use alternative routes to guide packet flow in ATM networks or even improve the efficiency of searching robots on the World Wide Web, in the sense of an increased multiconnectivity of its hyperlink structure.

Given a triple (G, s, t), where G is a $G_{n,p}$ random graph and s, t are two of its nodes, a natural notion of robustness is to require at least two edge disjoint

---

paths of short length (say, exactly ℓ or at most ℓ) between s and t, so that connectivity by short paths survives, even in the event of a link failure. We next give the following formal definition of $\ell - robustness$.

Definition 1 (ℓ-robustness). *A random graph G of the model $G_{n,p}$ is $\ell -$ robust for two vertices s, t when there are two edge-disjoint paths of length at most ℓ between s, t in G.*

In this work, we investigate the expected number of such paths between two vertices of the random graph, as well as bounds for the threshold probability $p_{n,\ell}$ (as functions of ℓ and n) for the existence of such paths in the $G_{n,p}$ random graph G.

Although $G_{n,p}$ has been extensively studied (see [1], [4], [14]), some questions of existence of multiple paths, which are vertex or edge disjoint between specific vertices have not been investigated till recently. The theory of random graphs began with the celebrated work of Erdös and Renyi ([7]) in 1959 and by now researchers know lot about the probable structure of these objects (see, e.g., the birth of the giant connected component in [11]). In this context we remark that the question of existence of many *vertex disjoint* paths of small length has been investigated by Nikoletseas et al in [13].

With respect to the corresponding question of the existence of many *edge* disjoint paths, we refer to the fundamental work of Broder et al ([5]). In that work, the authors show the existence of many edge disjoint paths in dense (with high probability connected) random graphs. However, the above work does not address the question of the existence of many edge disjoint paths *of a certain length*. Also, the estimation of the *precise number* of such paths (as a function of the density of the graph) still remains open.

It turns out that even the enumeration of paths among the vertices 1 and n that avoid all edges of the graph $(1, 2 \ldots, n)$ but pass through its vertices, is a non-trivial task. In fact, such an enumeration is a special case of enumerating permutations $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ of $(1, 2, \ldots, n)$ where certain gaps $\sigma_{i+1} - \sigma_i$ are forbidden. In our case, $\sigma_{i+1} - \sigma_i$ must not be in the set $\{-1, 1\}$.

In this work, we provide a precise estimate of the expected number of un-ordered pairs of paths in a random graph that connect a common source to a common destination, and have no edge in common, though they may share some nodes. Thus, for any given set of values of n, p, ℓ (where ℓ represents the length) we estimate precisely the mean number of avoiding pairs in graphs of a given size. This leads to a tight bound for the probability of non-existence of such paths between any fixed pair of nodes and also to a bound for the probability of the existence of many pairs of edge disjoint paths of length ℓ.

In order to achieve this, we devise a finite state mechanism that describes classes of permutations with free places and exceptions. The finite-state description allows for a direct construction of a multivariate generating function. The generating function is then subjected to an integral transform that implements an inclusion-exclusion argument and an explicit enumeration result; see Theorems 1 and 2. This enables us to quantify the trade-off between ℓ-robustness

(as defined above) and the density of the graph (i.e. the number of its edges). The originality of our approach consists in introducing in this range of problems methods of analytic combinatorics and recent research in automatic analysis (based on symbolic computation). For context, see [8], [9], [6]. Additional threshold estimates regarding properties of multiple source-destination pairs are discussed in the last two sections of the paper.

Summary of results: From earlier known results ([4], [13]) and this paper, a picture of robustness under the $G_{n,p}$ model emerges. (As usually in random graph theory, various regimes for $p = p(n)$ are considered). Start with an initially totally disconnected graph, corresponding to $p = 0$. As p increases, the graph becomes connected near the connectivity threshold $p_c(n) \simeq (\log n)/n$. All the following results hold for any integer ℓ: $6 \leq \ell < n$. Any fixed s, t pair (or equivalently a random s, t pair, given the invariance properties of $G_{n,p}$) *is "likely" to be ℓ-robust* when p crosses the value

$$p_{n,\ell}(n) = 2^{\frac{1}{2\ell}} \, n^{-1+\frac{1}{\ell}}$$

(see Theorem 3 and Equation 6). (Here "likely" means that the expected number of edge-disjoint pairs is at least 1 when $n \to \infty$). As long as $p \leq p_L(n, \ell)$, where

$$p_L(n, \ell) = n^{-1+\frac{1}{\ell}} \left(\log \frac{n^2}{\log n} \right)^{\frac{1}{\ell}}$$

we know, w.h.p., the *existence* of s, t pairs that are *not* connected by short (of length at most ℓ) paths; see Theorem 4. (The function $p_L(n)$ is in fact a threshold for diameter). However, one only needs a tiny bit more edges, namely $p \geq p_U$:

$$p_U(n, \ell) = n^{-1+\frac{1}{\ell}} \, 2 \left(\log \left(n^2 \log n \right) \right)^{\frac{1}{\ell}}$$

to ensure that *almost all* s, t-pairs are ℓ-robust; see Theorem 5.

2   The Enumeration of "Avoiding" Configurations

The problem at hand is that of estimating the expected number of "avoiding pairs" of length ℓ between a random source and a random destination in a $G_{n,p}$ random graph G. (An avoiding pair of length ℓ means an unordered pair of paths, each of length ℓ, that connect a common source to a common destination, and have no edge in common though they may share some nodes). This problem first necessitates the solution of enumeration problems that involve two major steps:

— Enumerate simple paths called "avoiding permutations" of length ℓ that can be viewed as hamiltonian paths on the set of nodes $[1, \ldots, \ell+1]$, connecting the source 1 and the destination $\ell + 1$, and having no edge of type $[i, i+1]$ or $[i, i-1]$.

— Enumerate so-called "avoiding paths", that are simple paths allowed to contain outer nodes taken from outside the segment $[1, \ldots, \ell+1]$. This situation is closer to the random graph problem since it allows nodes taken from the pool of vertices available in the graph $G \in G_{n,p}$.

The first problem is of independent combinatorial interest as it is equivalent to counting special permutations with restrictions on adjacent values. It also serves as a way to introduce the methods needed for the complete random graph problem. Both problems rely on the inclusion-exclusion principle that is familiar from combinatorial analysis and counting by generating functions (GF's). Applications of these results to the $G_{n,p}$ model are treated in the next section.

2.1   Symbolic Enumeration Methods

We use here a symbolic approach to combinatorial enumeration, according to which many general set-theoretic constructions have direct translations over generating functions. A specification language for elementary combinatorial objects is defined for this purpose. The problem of enumerating a class of combinatorial structures then simply reduces to finding a proper specification, a sort of a formal grammar, for the class in terms of basic constructions. (This general method has been carefully explained in the fundamental work of F. Chyzak, Ph. Flajolet and B. Salvy ([5])).

In this framework, classes of combinatorial structures are defined either iteratively or recursively in terms of simpler classes by means of a collection of elementary combinatorial constructions. The approach followed resembles the description of formal languages by means of context-free grammars, as well as the construction of structured data types in classical programming languages.

The approach developed here is direct, more "symbolic", as it relies on a specification language for combinatorial structures. It is based on so-called admissible constructions that have the important feature of admitting direct translations into generating functions. We specifically examine constructions whose natural translation is in terms of ordinary generating functions.

The ordinary generating function (OGF) of a sequence $\{A_n\}$ is, we recall, $A(z) = \sum_{n=0}^{\infty} A_n \cdot z^n$.

Definition 2 (Admissible Constructions). *Assume that Φ is a binary construction that associates to two classes of combinatorial structures B and C a new class $A = \Phi(B, C)$ in a finite way (each A_n depends on finitely many of the B_n and C_n). The Φ is admissible iff the counting sequence $\{A_n\}$ of A is a function of the counting sequences $\{B_n\}$ and $\{C_n\}$ of B and C only: $\{A_n\} = \Xi[\{B_n\}, \{C_n\}]$. In that case, there exists a well defined operator Ψ relating the corresponding ordinary generating functions: $A(z) = \Psi[B(z), C(z)]$.*

In this work, we will basically use three important constructions: union, product and sequence, which we describe below.

Definition 3 (Union Construction). *The disjoint union A of two classes B, C is the union (in the standard set-theoretic sense) of two disjoint copies, B^o and C^o, of B and C. Formally, we introduce two distinct "markers" ϵ_1 and ϵ_2, each of size zero, and define the (disjoint) union $A = B + C$ of B, C by $B + C = (\{\epsilon_1\} \times B) \cup (\{\epsilon_2\} \times C)$. The ordinary generating function is clearly $A(z) = B(z) + C(z)$. We represent the disjoint union construction by* Union.

Definition 4 (Product Construction). *If construction A is the cartesian product of classes B and C ($A = B \times C$), then, considering all possibilities, the counting sequences corresponding to A, B, C are related by the convolution relation: $A_n = \sum_{k=0}^{n} B_k \cdot C_{n-k}$ and the ordinary generating function is clearly $A(z) = B(z) \cdot C(z)$. We represent the product construction by* Prod.

Definition 5 (Sequence Construction). *If C is a class of combinatorial structures then the sequence class $\mathcal{G}\{\mathcal{C}\}$ is defined as the infinite sum $\mathcal{G}\{\mathcal{C}\} = \{\epsilon\} + \mathcal{C} + (\mathcal{C} \times \mathcal{C}) + \cdots$ with ϵ being a "null structure", meaning a structure of size 0. The null structure plays a role similar to that of the empty word in formal language theory and the sequence construction is analogous to the Kleene star operation (C^*). The ordinary generating function is clearly given by $A(z) = 1 + B(z) + B^2(z) + \cdots = \frac{1}{1-B(z)}$ where the geometric sum converges in the sense of formal power series since $[z^0]B(z) = 0$. We represent the sequence construction by* Sequence.

2.2   Avoiding Permutations

An avoiding permutation of length ℓ is a sequence $\tau = [\tau_1, \tau_2, \ldots, \tau_\ell, \tau_{\ell+1}]$ that is a permutation of $[1, \ldots, \ell + 1]$ and that satisfies the following conditions: $\tau_1 = 1$, $\tau_{\ell+1} = \ell + 1$, and $\tau_{i+1} - \tau_i \neq \pm 1$ for all i such that $1 \leq i \leq \ell$. Clearly, such a permutation encodes a path from 1 to $\ell + 1$ that has no edge in common with the graph $(1, 2, \ldots, \ell + 1)$. The parameter $\ell + 1$ is referred to as the size. There is no avoiding permutation for sizes $2, 3, 4, 5$. Surprisingly, the first nontrivial configurations occur at size 6, where the 2 possibilities are $[1, 4, 2, 5, 3, 6]$ and $[1, 3, 5, 2, 4, 6]$, while for size 7, there are 10 possibilities

$$[1, 3, 6, 4, 2, 5, 7], [1, 3, 5, 2, 6, 4, 7], [1, 4, 6, 2, 5, 3, 7], [1, 4, 2, 6, 3, 5, 7], [1, 5, 3, 6, 4, 2, 7],$$
$$[1, 5, 3, 6, 2, 4, 7], [1, 5, 2, 4, 6, 3, 7], [1, 4, 6, 3, 5, 2, 7], [1, 6, 3, 5, 2, 4, 7], [1, 6, 4, 2, 5, 3, 7].$$

The goal in this subsection is to determine the number Q_n of avoiding permutations of size n (that is, of length $n - 1$). We prove:

Theorem 1. *Avoiding permutations have ordinary generating function*

$$Q(z) := \sum_n Q_n z^n = \frac{\left(z - 1 + e^{\frac{1+z}{z(z-1)}} \, Ei\left(1, \frac{1+z}{z(z-1)}\right)\right) z}{(z-1)(1+z)} =$$

$$\frac{z \left(z + z \, hypergeom\left([1, 1], [\,], -\frac{z(z-1)}{1+z}\right) + 1\right)}{(1+z)^2}$$

where Ei *is the exponential integral and* hypergeom *represents the hypergeometric series. Equivalently, Q_n is expressible as a double binomial sum:*

$$Q_{n+2} = (-1)^{n-1} + \sum_{k_2=0}^{n} \sum_{k_1=0}^{n-k_2} (-1)^{k_1+k_2} (n-k_1-k_2)! \binom{n-k_1-k_2}{k_1} \binom{n+1-k_1}{k_2}$$

Proof. By the inclusion-exclusion principle, we need to determine the number $F_{n,j}$ of permutations $[\tau_1 = 1, \tau_2, \ldots, \tau_{n-1}, \tau_n = n]$, with at least j "exceptions", among which j distinguished, that are successions of values of the form $\tau_j - \tau_{j-1} = \pm 1$. The number of permutations with no exception is then:

$$Q_n = \sum_{j=0}^{n-1} (-1)^j F_{n,j} \tag{1}$$

A permutation with exceptions can be regarded as including a subcollection of "exceptional" edges that belong to the graph with edges $(1,2), (2,3), \ldots, (n-1, n)$. If we scan from left to right and group such exceptions by blocks, we get a *template*; a template thus represents a possible pattern of exceptional edges.

A template can be defined directly as made of blocks that are either: (*i*) isolated points; (*ii*) contiguous unit intervals oriented left to right (*LR*); (*iii*) contiguous unit intervals oriented right to left (*RL*). There is the additional constraint that the first and last blocks cannot be of type *RL*. For instance, for $n = 13$, the template $[[1, 2, 3], [4], [5, 6], [7], [8], [11, 10, 9], [12, 13]]$ will correspond to any permutation that has successions of values (in the cycle traversal) $1, 2; 2, 3; 5, 6; 11, 10; 10, 9; 12, 13$ as distinguished exceptions to the basic constraint of avoiding permutations.

We next provide the combinatorial specification for avoiding permutations.

Let $\{a, b\}$ be a binary alphabet. We now describe the grammar of templates.

The collection of strings beginning and ending with a letter a is described by the following rule:

sp0 := S = Prod(Sequence(Prod(a, Sequence(b))), a)

(It suffices to decompose according to each occurrence of the letter a). Now, the thre types of blocks in a template are described by the following rules:

sp1 := Prod(*begin_blockP, Z, end_blockP*)
sp2 := Prod(*begin_blockLR, Z*, Sequence(Prod(*mu_length, Z*), $1 \leq$ *card*), *end_blockLR*)
sp3 := Prod(*begin_blockRL*, Sequence(Prod(*mu_length, Z*), $1 \leq$ *card*), *Z, end_blockRL*)

Clearly, *sp2* and *sp3* are combinatorially isomorphic. For reasons related to the application of the inclusion-exclusion argument, we keep track of the size (number of nodes $l = 1$ of the basic interval graph, denoted by *card*) as well as of the length of blocks and of their *LR* or *RL* character, denoted by *mu_length*. Then, the grammar of templates is completed by substituting into *sp0*

$$a = \text{Union}(sp1, sp2) \quad \text{and} \quad b = sp3$$

Let $F_{n,k,j}$ be the number of templates with size n, k blocks and j exceptional edges. Then, counting the number of ways of linking blocks together, yields:

$$F_{n,j} = \sum_k F_{n,k,j}\, \phi(k) \tag{2}$$

where $\phi(k)$ is the Gamma integral:

$$\phi(k) \equiv (k-2)! = \int_0^\infty e^{-u}\, u^k\, \frac{du}{u^2}$$

for $k \geq 2$, and $\phi(1) = 1$ (since any such linking is determined by an arbitrary permutation of the $k - 2$ intermediate blocks). Observe that the extension of ϕ by linearity to an arbitrary series $\mathrm{h}(u)$ in u is given by

$$\phi(\mathrm{h}(u)) = \int_0^\infty \frac{e^{-u}\left(\mathrm{h}(u) - (u - u^2)\left(\frac{\partial}{\partial u}\mathrm{h}(u)\right)_{u=0}\right)}{u^2}\, du$$

That is to say, we just replace in expansions $u \to u^2$ and apply the Euler integral

$$\int_0^\infty e^{-u} u^k\, du = k!$$

Thus, with $\mathrm{F}(z, u, v) = \sum F_{n,k,l}\, z^n u^k v^l$, the OGF $Q(z) = \sum Q_n z^n$ satisfies

$$Q(z) = \phi(\mathrm{F}(z, u, -1))$$

Thus, from (1) and (2) above, everything boils down to obtaining the $F_{n,k,j}$.

Template enumeration. The approach to determining the sequence $F_{n,k,j}$ consists in introducing the trivariate GF, which immediately results from the above combinatorial specification.

$$F(z, u, v) = \sum_{n,k,\ell} F_{n,k,\ell}\, z^n u^k v^l$$

There, z records size, u records the total number of blocks (needed for subsequent permutation enumerations since blocks should be chained to each other), and v records the total length of LR or RL blocks (the number of distinguished exceptions needed for inclusion-exclusion).

We now carefully employ the generating functions for the union, product and sequence constructions in the grammar rules of the combinatorial specification for avoiding permutations defined above.

The set of words made of a's and b's that start and end with an a is described symbolically by

$$W = \frac{1}{1 - \frac{a}{1-b}} \cdot a \tag{3}$$

This is because $(1 - f)^{-1} = 1 + f + f^2 + f^3 + \cdots$ generates symbolically all sequences of objects of type f. Thus, W represents a sequence of objects of type

$\frac{a}{1-b}$ that start with an a. On the other hand, $\frac{a}{1-b}$ represents a sequence of objects of type b that end with an a. Take now the three types of blocks: isolated, LR, and RL. The GF's are, respectively, z, $LR(z) = z^2/(1-z)$, $RL(z) = z^2/(1-z)$. This is because isolated points are always of size 1, while LR and RL objects must be of size at least 2 (we have thus to multiply with z^2). Since the first and the last blocks can only be isolated points or LR blocks, the univariate GF for blocks is obtained by substituting a by $z + LR$ (isolated point or LR block) and b by RL in W. Thus we get the following trivariate GF:

$$F(z, u, v) = \left(1 - \frac{u\left(z + \frac{z^2 v}{1 - vz}\right)}{1 - \frac{uz^2 v}{1 - vz}}\right)^{-1} \cdot u\left(z + \frac{vz^2}{1 - vz}\right) =$$

$$= -\frac{uz(-1 + vz + uz^2 v)}{1 - 2vz - uz + v^2 z^2 + v^2 z^3 u}$$

Path counting. For the inclusion-exclusion argument, it is easy to observe that the desired sum $\sum_{n,k,\ell} F_{n,k,\ell} \, z^n u^k (-1)^\ell$ corresponds to the specialization $F(z, u, -1)$. This yields:

$$F(z, u, -1) = -\frac{uz(-1 - z - uz^2)}{1 + 2z - uz + z^2 + z^3 u} \tag{4}$$

Application of the ϕ-transformation (that counts the number of ways to connect the blocks) requires the modified form and so we get:

$$F(z, u, -1) = \frac{zu^2(uz^2 + 2z - uz + 1)}{(1 + z)(uz^2 + z - uz + 1)}$$

The corresponding ordinary generating function is

$$Q(z) = \int_0^\infty \frac{z(uz^2 + 2z - uz + 1)}{(1 + z)(uz^2 + z - uz + 1)} \cdot e^{-u} \, du$$

The quantity $Q(z)$ can be expressed in terms of the exponential integral

$$\int_0^\infty e^{-u} u^k \, du = k!$$

in the following closed form

$$Q(z) := \frac{\left(z - 1 + e^{\frac{1+z}{z(z-1)}} \operatorname{Ei}\left(1, \frac{1+z}{z(z-1)}\right)\right) z}{(z - 1)(1 + z)}$$

Since one deals with ordinary generating functions, this is to be taken as a formal (asymptotic) series. Note also that the exponential integral (Ei) involves the divergent series of factorials $\sum_{n=0}^\infty n!(-y)^{(-n-1)}$ which is also a hypergeometric series. This gives rise to a general conversion procedure from exponential

integrals to hypergeometric forms. Hence, another closed form for the OGF of the Q_n is

$$Q(z) := \frac{z \left(z + z \operatorname{hypergeom}\left([1, 1], [\,], -\frac{z(z-1)}{1+z}\right) + 1\right)}{(1 + z)^2}$$

Thus, with $F(z, u, v) = \sum F_{n,k,l} z^n u^k v^l$ and for OGF $Q(z) = \sum Q_n z^n$, by recalling that $Q(z) = \phi(F(z, u, -1))$ we get the expression for $Q(z)$ as stated in the Theorem. The expression can then be expanded using the binomial theorem, and double combinatorial sums result for coefficients. □

Though they have no direct bearing on the graph problem at hand, we mention two interesting consequences of this theorem.

Corollary 1. *The quantities Q_n satisfy the recurrence*

$$(n + 1)Q_n + Q_{n+1} - 2nQ_{n+2} + 4Q_{n+3} + (n + 3)Q_{n+4} - Q_{n+5} = 0,$$

where $Q(0) = 0$, $Q(1) = 1$, $Q(2) = Q(3) = Q(4) = Q(5) = 0$ and the asymptotic estimate

$$\frac{Q_n}{(n-2)!} = e^{-2}\left(1 + O\left(\frac{1}{n}\right)\right)$$

Proof. To get the recurrence relation, we use the following holonomic descriptions (introduced by Zeilberger), that is sequences that satisfy linear recurrences with polynomial coefficients:

$$\left(z^4 + z^5 + 4z^3 - 1 - z + 4z^2\right) Y(z) + \left(-2z^4 + z^2 + z^6\right) \left(\frac{\partial}{\partial z} Y(z)\right) -$$

$$-z^4 - 2z^3 \_C_0 - 2z^3 - z^4 \_C_0 + \_C_0 z^2 - z^5 + z - z^2 = 0$$

where $Y(z) = Q(z)$. By putting $C_0 = 1$, we get:

$$(z^4 + z^5 + 4z^3 - 1 - z + 4z^2)Y(z) + (-2z^4 + z^2 + z^6) \left(\frac{\partial}{\partial z} Y(z)\right) - 2z^4 - 4z^3 - z^5 + z = 0$$

We can now get (by elementary properties of the z-transform) the following transformation to a linear recurrence:

$$u(0) = 0, \ u(1) = 1, \ u(2) = 0, \ u(3) = 0, \ u(4) = 0, \ u(5) = 0,$$

$$(n+1)u(n) + u(n+1) - 2n \, u(n+2) + 4u(n+3) + (n+3)u(n+4) - u(n+5) = 0$$

We note that this provides an algorithm that uses a linear number of arithmetic operations to determine the quantities Q_n. By using the following principle based on the generating function method:

$$coeff_{z^n} \operatorname{hypergeom}\left([1, 1], [\,], z + dz^2 + O\left(z^3\right)\right) = n! \, e^d (1 + o(1))$$

provided that the argument of the hypergeometric is a function that is analytic at the origin, we have proved

$$Q(z) = \frac{z\left(z + z\,\text{hypergeom}\left([1,1],[\ ], -\frac{z(z-1)}{1+z}\right) + 1\right)}{(1+z)^2}$$

and since

$$-\frac{z(z-1)}{1+z} = z - 2z^2 + 2z^3 - 2z^4 + 2z^5 + O(z^6)$$

we have proved that the asymptotic proportion of legal permutations is *exactly* equal to e^{-2}.   □

The recurrence above implies the non-obvious fact that the number of avoiding permutations Q_n are computable in linear time. The asymptotic estimate extends properties known for permutations with excluded patterns (e.g., derangements have asymptotic density e^{-1}). Consequently, a nonzero proportion (about 13.53%) of all permutations that start with 1 and end with n are avoiding.

2.3   Avoiding Paths

We consider now the problem of counting the number $Q_{n,j}$ of avoiding paths of type (n,j), where n is the size (the number of nodes) and j is the number of "outer nodes". Such avoiding paths are defined by the fact that they satisfy the basic constraints of avoiding permutations regarding the base line $(1,2,\ldots,n)$, but contain in addition j outer nodes taken to be indistinguishable (unlabelled) and conventionally represented by the symbol '$\star$'. For instance, for types $(n,j) = (3,1),(4,1),(4,2)$, the listings are respectively

$$\{[1,\star,3]\} \qquad \{[1,3,\star,4],[1,\star,2,4]\} \qquad \{[1,\star,\star,4]\}$$

Theorem 2. *The number of avoiding paths is expressible as*

$$Q_{n+2,j} = \sum_{k_2=0}^{n-j}\sum_{k_1=0}^{n-j-k_2}(-1)^{k_1+k_2}(n-k_1-k_2)!\binom{n-j-k_1-k_2}{k_1}\cdot$$
$$\cdot\binom{n-j+1-k_1}{k_2}\binom{n-k_1-k_2}{j}^2.$$

(where $j \geq 0$)

Proof. We first define templates on which an inclusion-exclusion argument is applied. The specifications are a simple modification of the templates associated to avoiding permutations.

Let $\{a,b,x\}$ be a ternary alphabet. We now define the grammar of templates.

The collection of strings beginning with a and containing only one x that occurs at the end is described by the rule:

$$sp0 := S = \mathrm{Prod}(\mathrm{Sequence}(\mathrm{Prod}(a,\, \mathrm{Sequence}(b))),\, x)$$

(It suffices to decompose according to each occurrence of the letter a). We first need so-called "outer points" that are taken from outer space:

$$Outerpoints := \mathrm{Sequence}(\mathrm{Prod}(Z,\, mu\_outerpoint))$$

We also need "inner points":

$$Innerpoints := \mathrm{Sequence}(\mathrm{Prod}(Z,\, mu\_innerpoint))$$

Size is defined as the cumulative number of points in the pair of paths that underlies an avoiding path in the sense above: it is thus equal to the length of the avoiding path plus the number of $\star$ symbols corresponding to the outer nodes. We thus introduce a special notation for nodes of the integer line that are shared by the two paths:

$$Z2 := \mathrm{Prod}(Z, Z)$$

Now, the three types of blocks are described by the following rules:

$sp1 := \mathrm{Prod}(mu\_block,\ Z2,\ Outerpoints,\ Innerpoints)$
$sp2 := \mathrm{Prod}(mu\_block,\ Z2,\ \mathrm{Sequence}(\mathrm{Prod}(mu\_length,\ Z2)),\ \mathrm{card} \geq 1),$
$$Outerpoints,\ Innerpoints)$$
$sp3 := \mathrm{Prod}(\mathrm{Sequence}(\mathrm{Prod}(mu\_length,\ Z2)),\ \mathrm{card} \geq 1),\ Z2,\ Outerpoints,$
$$Innerpoints,\ mu\_block)$$

(Clearly, $sp2$ and $sp3$ are combinatorially isomorphic). The blocks that can occur at the end are of type x and can only be of type $sp1$ or $sp2$ but without outer points nor inner points.

$$sp1x := \mathrm{Prod}(mu\_block,\ Z2)$$
$$sp2x := \mathrm{Prod}(mu\_block,\ Z2,\ \mathrm{Sequence}(\mathrm{Prod}(mu\_length,\ Z2)),\ 1 \leq \mathrm{card}))$$

The above grammar is completed (to give S) by substituting into $sp0$

$$a = \mathrm{Union}(sp1, sp2)$$
$$b = sp3 \quad \text{and}$$
$$x = \mathrm{Union}(sp1x, sp2x)$$

The 5-variate GF immediately results from the above specification:

$$\mathrm{F}(z, u, v, w_1, w_2) :=$$

$$\frac{-u\, z^2 (-1 + z\, w_2 + z\, w_1 - z^2\, w_1\, w_2 + v\, z^2 - v\, z^3\, w_2 - v\, z^3\, w_1 + v\, z^4\, w_1\, w_2 + u\, z^4\, v)}{1 - z(w_2 + w_1) + z^2(w_1\, w_2 - u) + v\, z^2(-2 + 2z\, w_1 - 2\, z^2\, w_1\, w_2 + v\, z^2(1 - z\, w_2 - z\, w_1 + z^2\, w_1\, w_2 + z^2\, u))}$$

where u, v, w_1, w_2 represent the blocks, the length, the outer nodes and the inner nodes, respectively.

For inclusion-exclusion, we set $v = -1$. Application of the ϕ-transformation (that counts the number of ways to connect the blocks) requires the modified form

$$F(z, u, -1, w_1, w_2) :=$$

$$\frac{u^2 \, z^2 \, (1 + 2z^2 - z^3 \, w_1 + u \, z^4 + z^4 \, w_1 \, w_2 - z^3 \, w_2 - u \, z^2 - z \, w_2 - z \, w_1 + z^2 \, w_1 \, w_2)}{(1 + z^2)(z^4 \, w_1 \, w_2 + u \, z^4 - z^3 \, w_2 - z^3 \, w_1 + z^2 \, w_1 \, w_2 + z^2 - u \, z^2 - z \, w_2 - z \, w_1 + 1)}$$

The ordinary generating function is here

$$Q(z) := \int_0^\infty \frac{z^2 (1 + 2z^2 - z^3 \, w_1 + u \, z^4 + z^4 \, w_1 \, w_2 - z^3 \, w_2 - u \, z^2 - z \, w_2 - z \, w_1 + z^2 \, w_1 \, w_2) \, e^{-u}}{(1 + z^2)(z^4 \, w_1 \, w_2 + u \, z^4 - z^3 \, w_2 - z^3 \, w_1 + z^2 \, w_1 \, w_2 + z^2 - u \, z^2 - z \, w_2 - z \, w_1 + 1)} \, du$$

And this can be expressed in terms of the exponential integral as follows:

$$Q(z) := z^2 \cdot e^{-\frac{(1+z^2)(w_2+w_1)}{z(z-1)(1+z)}}$$

$$\cdot \frac{\mathrm{Ei}\left(1, \dfrac{z^4 \, w_1 \, w_2 - z^3 \, w_2 - z^3 \, w_1 + z^2 \, w_1 \, w_2 + z^2 - z \, w_2 - z \, w_1 + 1}{z^2 (z^2 - 1)}\right) e^{\frac{(1+z^2)(z^2 \, w_1 \, w_2 + 1)}{z^2(1-z^2)}}}{(z^2 - 1)(1 + z^2)} +$$

$$+ z^2 \cdot e^{-\frac{(1+z^2)(w_2+w_1)}{z(z-1)(1+z)}} \cdot \frac{z^2 e^{\frac{(1+z^2)(w_2+w_1)}{z(1-z^2)}} - e^{\frac{(1+z^2)(w_2+w_1)}{z(1-z^2)}}}{(z^2 - 1)(1 + z^2)}$$

Again, there is an "explicit form" of the OGF of the problem

$$Q(z) := z^2 \cdot$$

$$\cdot \frac{z^3 (z \, w_1 \, w_2 - w_2 - w_1) + z^2 \, \mathrm{hypergeom}\left([1, 1], [\,], -\dfrac{z^2 \, (z-1)(1+z)}{(1+z^2)(z \, w_2 - 1)(z \, w_1 - 1)}\right)}{(1 + z^2)^2 (z \, w_2 - 1)(z \, w_1 - 1)} +$$

$$+ z^2 \cdot \frac{z^2 (1 + w_1 \, w_2) - z(w_2 + w_1) + 1}{(1 + z^2)^2 (z \, w_2 - 1)(z \, w_1 - 1)}$$

and also

$$Q(z) := \frac{z^4 \, \mathrm{hypergeom}\left([1, 1], [\,], -\dfrac{z^2 \, (z-1)(1+z)}{(1+z^2)(z \, w_2 - 1)(z \, w_1 - 1)}\right)}{(1 + z^2)^2 (z \, w_2 - 1)(z \, w_1 - 1)} + \frac{z^2}{1 + z^2}$$

The coefficient $c(n, j, k)$ of $z^n \, w_1^{\,j} \, w_2^{\,k}$ is obtained by straight expansion and avoiding paths are then enumerated by $C(n, j) = c(n, j, j)$. The corresponding formulae of the Theorem statement are obtained directly by symbolic expansions.

The computations are rather intensive and, for instance, the 4-variable GF that "lifts" $F(z, u, -1)$ is found to be

$$\frac{u \, z^2 \left(1 - z \, w_2 - z \, w_1 + z^2 \, w_1 \, w_2 + z^2 - z^3 \, w_2 - z^3 \, w_1 + z^4 \, w_1 \, w_2 + u \, z^4\right)}{(1 + z^2) \left(z^4 \, w_1 \, w_2 + u \, z^4 - z^3 \, w_2 - z^3 \, w_1 + z^2 \, w_1 \, w_2 + z^2 - u \, z^2 - z \, w_2 - z \, w_1 + 1\right)} \tag{5}$$

It is to be noted that computations have been performed with the help of the computer algebra packages `Combstruct` and `Gfun` that are dedicated to automating computations in combinatorial analysis and have been developed in the Maple system for symbolic computation.   □

3   Average-Case Analysis for the Random Graph Model

We show now how to estimate the robustness to link failures in a random graph that obeys the $G_{n,p}$ model. An *avoiding pair* of length ℓ in a graph is an *unordered* pair of paths, each of length ℓ, with a common source and a common destination, that may share some nodes, but are edge disjoint. We have:

Theorem 3. *The mean number of avoiding pairs of length ℓ between a random source and a random destination in a random graph obeying the $G_{n,p}$ model is*

$$N_\ell(n,p) := \frac{p^{2\ell}}{2n(n-1)} \sum_{j=0}^{\ell} Q_{\ell+1,j} \binom{n}{l+1+j} (l+1+j)!$$

where the coefficients $Q_{n,j}$ are given by Theorem 2.

Proof. The coefficient $1/2$ corresponds to the fact that one takes unordered pairs of paths; the coefficient $1/(n(n-1))$ averages over all possible sources and destinations; the factor $p^{2\ell}$ provides the edge weighting corresponding to $G_{n,p}$; the arrangement numbers account for the number of ways to embed an avoiding path into a graph by choosing certain nodes and assigning them in some order to an avoiding path; the coefficients $Q_{\ell+1,j}$ provide the basic counting of avoiding paths that build up avoiding pairs. □

Note 1. Since the $G_{n,p}$ model implies isotropy, the quantity $N_\ell(n,p)$ is also the mean number of avoiding pairs between *any fixed* source and destination s, t.

Robustness. A short table of initial values of $N_\ell(n,p)$ follows:

$$N_2 = \tfrac{1}{2}(n-2)(n-3)p^4 \qquad N_3 = \tfrac{1}{2}(n-2)(n-3)^2(n-4)p^6$$
$$N_4 = \tfrac{1}{2}(n-1)(n-2)(n-3)(n-4)(n-5)^2 p^8$$
$$N_5 = \tfrac{1}{2}(n-2)(n-3)(n-4)(n-5)^2(n^3 - 11n^2 + 25n + 32)p^{10}$$

From developments in the previous section, the formulæ are computable in low polynomial time (as a function of ℓ). They make it possible to estimate the mean number of avoiding pairs in graphs of a given size for all reasonable values of n, p, ℓ. Take for instance a graph with $n = 10^5$ nodes and an edge probability $p = 5 \cdot 10^{-5}$. This corresponds to a mean node degree that is extremely close to 5, so that, on average, each node has 5 neighbors. Then the mean values are

$N_2 = 3.1 \cdot 10^{-6}$, $N_3 = 7.8 \cdot 10^{-5}$, $N_4 = 1.9 \cdot 10^{-5}$, $N_5 = 4.8 \cdot 10^{-4}$, $N_6 = 1.2 \cdot 10^{-2}$, $N_7 = 0.30$, $N_8 = 7.6$, $N_9 = 190$, $N_{10} = 4763$, $N_{11} = 119062$, $N_{12} = 2.9 \cdot 10^7$

Thus, in this example, one expects to have short and multiple connections between source and destination provided paths of length 8 are allowed. This numerical example also shows that there are rather sharp transitions. The formula of Theorem 3, that entails the following rough approximation

$$N_\ell(n,p) \approx \frac{1}{2} \, n^{2\ell-2} p^{2\ell} \tag{6}$$

precisely accounts for such a sharpness phenomenon.

In the introduction, we have defined ℓ-robustness as multiple connectivity by edge-disjoint paths of length *at most* ℓ. In fact, Equation 5 leads to explicit expressions for generalized avoiding pairs of type (ℓ_1, ℓ_2) that are made of two paths, of lengths ℓ_1, ℓ_2. It can then be seen that the bottleneck for existence of pairs (ℓ_1, ℓ_2) with $\ell_1, \ell_2 \leq \ell$ is in fact the case (ℓ, ℓ). Thus, since $N_\ell(n, p) \to 0$ when $\frac{p}{p_r(n,\ell) \to 0,}$ the function

$$p_r(n, \ell) = 2^{\frac{1}{2\ell}} n^{-1+\frac{1}{\ell}}$$

is a cut-off point for ℓ-robustness and an $(\leq \ell, \leq \ell)$-avoiding pair is expected or not depending on whether p/p_r tends to 0 or to ∞.

Corollary 2. *Any fixed pair in a $G_{n,p}$ graph is almost certainly not $\ell - robust$ if $p/p_r(n) \to 0$.*

Proof. When $\frac{p}{p_r(n,\ell)} \to 0$, then the expected number $N_\ell(n, p)$ of the desired pairs of paths tends to 0 and so does the probability of existence of at least one such pair of paths (since this probability, by Markov Inequality, is bounded from above by the expectation). Thus, with probability tending to 1, there is no pair of edge disjoint paths between the two vertices and these two vertices are, almost certainly, not $\ell - robust$. $\square$

4   Thresholds in the Random Graph Model

In this section we provide bounds for the probability (and thus the threshold, if it exists) of the existence, between any fixed pair of vertices, of two edge-disjoint paths of length at most ℓ, by proving the following:

- We give an estimation of the value $p_L \equiv p_L(n, \ell)$ such that $G_{n,p}$ graphs with $p \leq p_L$ do not satisfy the desired property of the existence, between any fixed pair of nodes, of two edge-disjoint paths between some pair of vertices, with probability tending to 1 as n goes to infinity.
- We present a value $p_U \equiv p_U(n, \ell)$ such that almost every $G_{n,p}$ graph with $p \geq p_U$ has almost all its source-destination pairs of vertices connected by at least two edge-disjoint paths of length at most ℓ.

Theorem 4. *Define*

$$P_L(n, \ell) = \left(\log \frac{n^2}{\log n} \right)^{\frac{1}{\ell}} n^{-1+\frac{1}{\ell}}$$

Then, for $p \leq P_L(n, \ell)$, almost surely, there exists a pair of vertices in the $G_{n,p}$ graph that does not have the $\ell - robustness$ property.

Proof. By using the threshold function for diameter. $\square$

Theorem 5. *Define*

$$P_U(n, \ell) = 2 \left(\log \left(n^2 \log n \right) \right)^{\frac{1}{\ell}} n^{-1+\frac{1}{\ell}}$$

Then, for $p \geq P_U(n, \ell)$, almost surely, almost all pairs of vertices of a $G_{n,p}$ graph have the $\ell - robustness$ property.

Proof. Consider two independent distributions G_{n,p_1} and G_{n,p_2} on the same set of vertices. Let $E_i (i = 1, 2)$ be the events "G_{n,p_i} has diameter ℓ".

Consider the graph $\tilde{G}$ obtained when we superimpose an instance $G' \in G_{n,p_1}$ and an instance $G'' \in G_{n,p_2}$ and OR them (i.e., $\tilde{G}$ has an edge joining u, v iff at least one of G', G'' has). Clearly $\tilde{G} \in G_{n,p}$ with

$$p = p_1(1 - p_2) + p_2(1 - p_1) + p_1 p_2 = p_1 + p_2 - p_1 p_2$$

In fact, if u, v are joined in G' by a path P_1 and in G'' by a path P_2, then these two paths both exist in $\tilde{G}$. For p around the threshold for diameter ℓ of $G_{n,p}$ and $\ell = o(n)$, the number of pairs u, v of $\tilde{G}$ for which the paths of G', G'' overlap is $o(n^2)$, thus the vast majority of pairs of vertices $(n^2 - o(n^2)$ of them) in $\tilde{G}$ is connected via two edge disjoint paths of length $\leq \ell$.

This gives approximately a $p_U \leq p_1 + p_2 - p_1 p_2$ and if $p_1 = p_2 = p_0^{(\ell)}$ (p_0 a threshold for diameter ℓ or $\ell + 1$) then

$$p_U \leq 2p_0^{(\ell)} - \left(p_0^{(\ell)} \right)^2 \leq 2(2 \log n - \log c)^{\frac{1}{\ell}} n^{\frac{1}{\ell}-1}$$

from [4], where c can be adjusted so that the diameter is almost surely ℓ (see [4], Corollary 12, p. 237). $\qquad\square$

5   A Discussion on the Extension to All Pairs

In this section we show how our results can be used to provide a tighter bound for p_U.

Lemma 1. *For every graph $G(V, E)$, if vertices u, v are each connected to a specific vertex $x \in V$ via two edge disjoint paths each of length ℓ, then u, v are connected in G via two edge disjoint paths, each of length at most 3ℓ.*

Proof. For simplicity, let the two (edge disjoint) paths from u to x be colored blue and the two (edge disjoint) paths from v to x be colored red. Take one of the two red paths and mark the first red-blue intersection vertex x_1 of it (it always exists such a vertex since it is in the worst case $x_1 = x$). Now take the other red path and mark the first red-blue intersection vertex x_2 (again this vertex can be x). There are two cases:
Case 1: Vertices x_1, x_2 are in different blue paths. Then the Lemma is easily proved by simply following the two different blue parts and then continuing with the two different red ones. Note that the two blue parts are edge disjoint,

the two red continuations are also edge disjoint and there is no red-blue edge.

Case 2: Both x_1, x_2 are on the same blue path. Let x_1 the closest to u on this blue path. Take the first $u - v$ path to be from u (on this blue path) to x_1 and then from x_1 to v (by the same red path which defined x_1) and the second $u - v$ path be composed by the other red path from v to x_2, then the blue part from x_2 to x and then the unused other blue path returning to u. Again, there is obviously no edge intersection.

With respect to length, the worst case is clearly Case 2, where the second constructed path has pieces from three of the four initial paths, leading to length at most 3ℓ. □

Lemma 1 yields the following corollary:

Corollary 3. *For every graph $G(V, E)$ if there exists a vertex $x \in V$ such that for all vertices $u, v \in V (u, v \neq x)$ each of u, v connects to x via two edge disjoint paths of length at most ℓ, then the diameter of G is at most 3ℓ and each $u, v \in V$ is connected via two edge disjoint paths of length at most 3ℓ.*

Theorem 6. *Given $G_{n,p}$, if $p(n, \ell)$ is such that the probability that two specific nodes of G are connected via two edge disjoint paths of length at most ℓ is at least $1 - \theta$ (where $\theta = o\left(\frac{1}{n}\right)$), then all pairs of nodes u, v of G are each connected via two edge disjoint paths of length at most 3ℓ with probability at least $1 - n\theta$.*

Proof. By applying to the probability of existence of paths between all pairs of vertices the fact that the probability of a union of events (existence of paths between two specific nodes) is bounded from above by the sum of the probabilities. □

Theorem 6 can provide an upper bound for the all pairs problem, by using an upper bound p_u such that for $p \geq p_u$, for every instance of $G_{n,p}$, any fixed (or random) pair has the $\ell - robustness$ property with probability tending to 1 as n tends to infinity. The derivation of such a bound could be approached by the computation of the Second Moment of the $\ell - robustness$ distribution, a computation that seems to be of major technical difficulty, and will be further examined in the future.

6   Conclusions and Further Research

We estimated here tightly and also asymptotically the mean value of the ways to get at least two edge disjoint paths between any two specific nodes of $G_{n,p}$ graphs. We pose as an open problem the calculation of the second moment for strengthening the threshold and also the extension of the problem to the existence of k edge disjoint paths.

References

1. N. Alon and J. Spencer, "The Probabilistic Method", John Wiley & Sons, 1992.
2. D. Bauer, F. Boesch, C. Suffel and R. Tindell, "Connectivity extremal problems and the design of reliable probabilistic networks", The Theory of Applications of Graphs, John Wiley and Sons, 1981.
3. F. T. Boesch, F. Harary and J. A. Cabell, "Graphs and models of communication networks vulnerability: connectivity and persistence", Networks, vol. 11, pp. 57–63, 1981.
4. B. Bollobás, "Random Graphs", Academic Press, 1985.
5. A. Broder, A. Frieze, S. Suen and E. Upfal, "Optimal construction of edge-disjoint paths in random graphs", SIAM Journal on Computing 28, 541-574. Also in the Proceedings of the 5th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 603 - 612, 1994.
6. F. Chyzak, Ph. Flajolet and B. Salvy, "Studies in automatic Combinatorics, Volume II", INRIA 1998. Available at URL
http://pauillac.inria.fr/algo/libraries/autocomb/autocomb.html
7. P. Erdös and A. Renyi, "On the evolution of random graphs", Magyar Tud. Akad. Math. Kut. Int. Kozl. 5, pp. 17–61, 1960.
8. Ph. Flajolet and R. Sedgewick, "An Introduction to the Analysis of Algorithms", Addison Wesley, 1996.
9. Ph. Flajolet and R. Sedgewick, "Analytic Combinatorics", book in preparation, 1999 (Individual chapters are available as INRIA Research Reports 1888, 2026, 2376, 2956, 3162).
10. J. Hromkovic, R. Klasing, E. Stoehr and H. Wagener, "Gossiping in Vertex-Disjoint Paths Mode in d-dimensional Grids and Planar Graphs", Proceedings of the 1st European Symposium on Algorithms (ESA), pp. 200-211, LNCS vol. 726, 1993.
11. S. Janson, D. Knuth, T. Łuczak and B. Pittel, "The Birth of the Giant Component", Random Structures and Algorithms, vol. 4, pp. 232-355, 1993.
12. Z. Kedem, K. Palem, P. Spirakis and M. Yung, "Faulty Random Graphs: reliable efficient-on-the-average network computing", Computer Technology Institute (CTI) Technical Report, 1993.
13. S. Nikoletseas, K. Palem, P. Spirakis, M. Young, "Short Vertex Disjoint Paths and Multiconnectivity in Random Graphs: Reliable Network Computing", 21st International Colloquium on Automata, Languages and Programming (ICALP), Jerusalem, pp. 508 – 515, 1994.
14. J. Spencer, "Ten Lectures on the Probabilistic Method", SIAM, 1987.