HAL
open science

Distances in random Apollonian network structures

Olivier Bodini, Alexis Darrasse, Michèle Soria

To cite this version:

Olivier Bodini, Alexis Darrasse, Michèle Soria. Distances in random Apollonian network structures. 20th Annual International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2008), Jun 2008, Viña del Mar, Chile. pp.307-318, 10.46298/dmtcs.3641 . hal-00345749

HAL Id: hal-00345749

https://inria.hal.science/hal-00345749v1

Submitted on 19 Aug 2015

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

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

FPSAC 2008, Valparaiso-Viña del Mar, Chile DMTCS proc. **AJ**, 2008, 307–318

Distances in random Apollonian network structures†

Olivier Bodini, Alexis Darrasse, and Michèle Soria

LIP6 - Equipe SPIRAL, 104 avenue du président Kennedy, F-75016 Paris, France
Université Paris 6 - UPMC, F-75252 Paris cedex 05

Abstract. In this paper, we study the distribution of distances in random Apollonian network structures (RANS), a family of graphs which has a one-to-one correspondence with planar ternary trees. Using multivariate generating functions that express all information on distances, and singularity analysis for evaluating the coefficients of these functions, we prove a Rayleigh limit distribution for distances to an outermost vertex, and show that the average value of the distance between any pair of vertices in a RANS of order n is asymptotically $\sqrt{n}$.

Résumé. Nous étudions dans ce papier la distribution des distances dans les structures des réseaux apolloniens aléatoires (RANS), une famille de graphes en bijection avec les arbres ternaires planaires. En s'appuyant sur l'utilisation de séries génératrices multivariées pour décrire toute l'information sur les distances, ainsi que sur l'analyse de singularités pour évaluer les coefficients de ces séries, nous prouvons une distribution limite de Rayleigh pour les distances vers un sommet externe du RANS et montrons que la distance moyenne entre deux sommets quelconques d'un RANS d'ordre n est asymptotiquement $\sqrt{n}$.

Keywords: formal power series, random networks, distance, singularity analysis.

1 Introduction

Many graph models have been recently introduced for representing the structure and dynamics of real-life networks (see e.g. [3]). Their adequacy to data can be measured by comparing some properties of graphs, especially the degree distribution of the vertices, which is related to scale-free properties, and properties related to the "small world" effect, such as distance between pairs of vertices and grouping in clusters.

The random Apollonian networks (RAN) proposed by [4] provides an interesting model, with a power-law degree distribution, a mean distance of logarithmic order and a large clustering coefficient. We introduced in [1] a modified version, random Apollonian network structures (RANS), which preserves the interesting properties of real graphs concerning degree distribution (a power-law with an exponential cut-off) and large clustering. This paper is devoted to the analysis of distances in RANS, which is showed to be of square root order. This result may seem disappointing in the scope of "small world" networks, however one only knows that the distance in real networks is small, not necessarily logarithmic.

†work supported by ANR contract GAMMA, n°BLAN07-2_195422

1365–8050 © 2008 Discrete Mathematics and Theoretical Computer Science (DMTCS), Nancy, France

Anyhow, there is no surprise with the square root distance since our underlying model is simple families of trees, whereas the model underlying RAN is increasing trees (that lead to a logarithmic distance [5]). Relying on these different tree models and using probabilistic methods, a unified study of various parameters in RAN and RANS was concurrently done [6].

A RANS can indeed be seen as a certain type of triangulation of a triangle, and the study of RANS relies on the bijection with planar ternary trees (see figure 1). From this bijection we can express the enumerative generating function for RANS, and use multivariate functions for marking several distance parameters. Moreover the asymptotic values of the quantities under consideration can be dealt with using singularity analysis (according to methods developed in [2]).

We are interested in two types of parameters measuring distance, and develop two methods to handle them. We first attack the distances between a special vertex (an outermost vertex A of the RANS) and all the other vertices. The method is built on computing a series of generating functions with increasingly many variables, that contains all informations concerning distances from A to the other vertices. Distribution analysis is based on the study of partial derivatives of this multivariate series, which correspond to the series counting the number of vertices at a certain distance from A. These series all express in terms of the generating function for RANS and asymptotic analysis gives a Rayleigh limit distribution with a mean value of order $\sqrt{3\pi n}/11$.

The second study addresses the total distance between all pairs of vertices. We exhibit a generating function in four variables that expresses simultaneously distances from one, two or three outermost vertices. This generating function has a nice recursive definition, due to the symmetries of the problem. It contains all information to compute the total distance between pairs of vertices. Geometrical considerations splits this total distance in two parts, depending on whether a path between two vertices spans over disjoints sub-RANS or not. The resulting mean distance between two vertices is of order $\sqrt{3\pi n}/11$.

This paper divides in four sections: this introduction, followed by a section that recalls the definition of random Apollonian network structures, the bijection with ternary trees, and the result for degree distribution. Section 3 describes the distribution of distances from an outermost vertex and section 4 is dedicated to the study of the total distance between all pairs of vertices.

2 Random Apollonian network structures

The recursive definition of RANS shows a one-to-one correspondence with ternary trees. The degree distribution, which is a power law with an exponential cut-off, was studied in [1] by considering bivariate series marking the corresponding parameter in trees.

2.1 Bijection with ternary trees

A random Apollonian network structure (RANS) R is recursively defined as: either an empty triangle or a triangle T split in three parts, by placing a vertex v inside T and connect it to the three vertices of the triangle; each sub-triangle being substituted by a RANS (see figure 1).

The vertices of T will be called the *outermost vertices* of R (noted $\mathcal{O}(R)$); and vertex v will be called the *center* of R. We will note $\mathcal{R}$ the class of all RANS.

The *order* of the empty RANS is zero and the order of a non-empty RANS is one plus the sum of the orders of the three sub-RANS.

Proposition 2.1 *[1] There is a bijection between random Apollonian network structures of order N and rooted plane ternary trees of size N (with N internal nodes).*

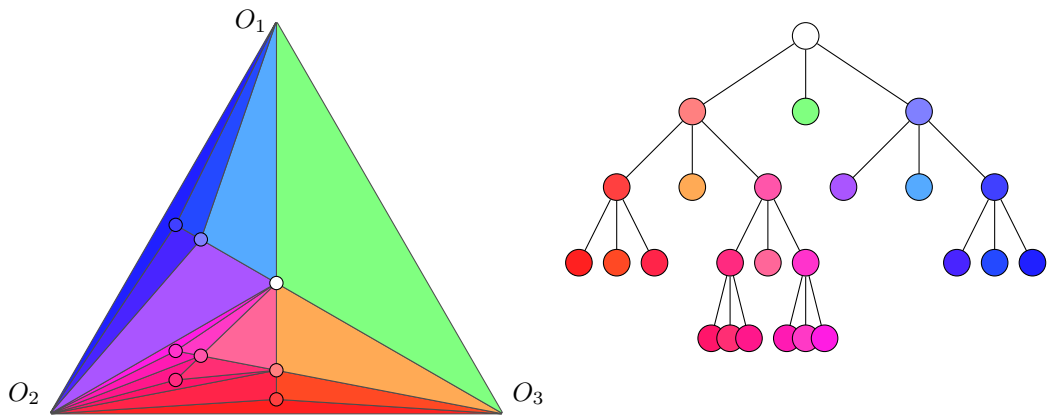

Fig. 1: A random Apollonian network and its corresponding ternary tree

In planar ternary trees, the linear ordering of siblings is relevant. This order is carried over to triangles: naming $\{O_1, O_2, O_3\}$ the vertices of $\mathcal{O}(R)$, imposes a linear ordering on the sub-RANS ($\{S_1, S_2, S_3\} = \mathcal{S}(R)$): S_i will be the one not containing O_i. Recursively replacing the missing outermost vertex by the center of R preserves the order in sub-RANS.

The generating function for ternary trees $T(z) = \sum T_N z^N$, where T_N is the number of trees with N internal nodes, satisfies the functional equation $T(z) = 1 + zT^3(z)$. $T(z)$ has radius of convergence $\rho = 4/27$ and singular value $\tau = 3/2$; and the singular expansion of $T(z)$ near ρ is

$$T(z) = \frac{3}{2} - \frac{\sqrt{3}}{2}\sqrt{1 - z/\rho} + \frac{2}{3}(1 - z/\rho) - \frac{35\sqrt{3}}{108}(1 - z/\rho)^{3/2} + O\left((1 - z/\rho)^{5/2}\right). \tag{1}$$

Thus the asymptotic form of the coefficients: $T_N \sim c\rho^{-N}N^{-3/2}$, with $c = \sqrt{3}/4\sqrt{\pi}$.

The derivative $T'(z) = \frac{T^3(z)}{1-3zT^2(z)}$ will also appear in the computations below. The leading term in its singular expansion is $\frac{\sqrt{3}}{4\rho}(1 - z/\rho)^{-1/2}$, thus a coefficient T'_N of asymptotic order $\frac{27\sqrt{3}}{16\sqrt{\pi}}\rho^{-N}N^{-1/2}$.

2.2 Degree distribution

The degree distribution in random Apollonian network structures follows a power law with an exponential cutoff. This is obtained [1] by analysing the degree of the center of a RANS, and propagating this study to the whole of the sub-RANS.

The bivariate generating function marking the degree of the center is $D_g(z, u) = zu^3T^3(z, u)$, where $T(z, u)$ is the bivariate generating function for ternary trees with u marking the degree of an outermost vertex in the corresponding RANS:

$$T(z, u) = 1 + uzT(z)T^2(z, u). \tag{2}$$

The degree of a vertex is exactly the number of vertices at distance 1 from this vertex, the generating functions used in the following section will be generalizations of $T(z, u)$.

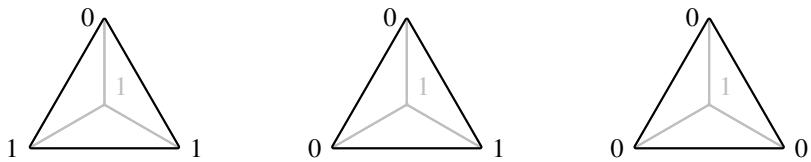

Fig. 2: RANS of type $(0,1,1)$, $(0,0,1)$, $(0,0,0)$ and their sub-RANS.

3 Distance from an outermost vertex

This section is devoted to computing the distribution of the distances from a fixed specific outermost vertex. We introduce a generating function with infinitely many variables, each variable u_i marking the number of vertices at distance i from the outermost vertex. Relying on the symmetries of the problem and the recursive nature of RANS, we are able to express and study this generating function.

The interest of this analysis is not only to find, with a different method, some results of the following section; but moreover this study can be adapted to compute the distribution of distances to *any distinguished vertex* in a RANS, which may be considered as a more realistic parameter.

3.1 Multivariate generating function

Due to the recursive nature of RANS, we often have to consider a RANS R as having an *environment*, that is a bigger RANS containing R. Given a RANS R, the distance of any of its vertices to a vertex v in the environment of R is determined by the three distances of the elements of $\mathcal{O}(R)$ to v.

Since the outermost vertices of R form a clique, their distances to any vertex cannot differ by more than one. This observation allows us to reduce our study to a few cases. First we work modulo a translation and restrict ourselves to the case when the three distances to $\mathcal{O}(R)$ are either 1 or 0. Second we can work modulo a permutation and restrict ourselves to only three cases: $(0,1,1)$, $(0,0,1)$ and $(0,0,0)$, illustrated in figure 2. These three cases actually correspond to labelling the internal vertices of R by their distances either to one (out of three), or to two (out of three) or to all three outermost vertices.

Definition The $\delta_{\mathcal{O}}$*-labeling* of $R \in \mathcal{R}$ consists in putting on each vertex a label corresponding to its distance from $O_1(R)$ (or equivalently to one of any $O_i(R)$):

- the outermost vertices O_1, O_2, O_3 respectively receive labels 0, 1 and 1;
- the center of $R' \in \mathcal{S}(R)$ is labeled by 1 plus the minimum of the labels of $\mathcal{O}(R')$.

Definition Let us define the *type of* $R \in \mathcal{R}$ as the 3-tuple of labels of $\mathcal{O}(R)$. We say that

- two RANS types are *equivalent* iff they have the same type up to a permutation.
- two RANS types are *translated* by θ iff their labellings are the same up to a translation θ.

In a RANS R of type $(0,1,1)$ the center gets label 1. Thus R is equivalent to $S_2(R)$ and $S_3(R)$, but it is not equivalent to $S_1(R)$, which is of type $(1,1,1)$: if a vertex is at distance d from $O_1(S_1(R))$, its distance from $O_1(R)$ is $d+1$.

This remark leads to a recursive definition of $T_1(z,u_1) = \sum T_{n,k} u_1^k z^n$, where $T_{n,k}$ is the number of RANS of order n with k vertices at distance 1 from $O_1(R)$:

$$T_1(z,u_1) = 1 + z u_1 T_1^2(z,u_1) T(z). \tag{3}$$

This equation follows the recursive definition of RANS, noticing first that the center is at distance 1 from $O_1(R)$ and second that the configuration is the same in $S_2(R)$ and $S_3(R)$, whereas in $S_1(R)$ there is no vertex at distance 1 from $O_1(R)$.

The problem of marking *both* vertices at distance 1 and 2 from $O_1(R)$ is treated in the same way: the configuration of R, of type $(0,1,1)$, recursively occurs in $S_2(R)$ and $S_3(R)$, which are of the same type. But the case of $S_1(R)$, of type $(1,1,1)$, is a little more tricky and requires a deeper decomposition. If $S_1(R)$ is not empty, its center is at distance 2 from $O_1(R)$, and its three sub-RANS $\mathcal{S}(S_1(R))$ are equivalent, of type $(1,1,2)$: either they are empty, or their center is at distance 2 from $O_1(R)$, one sub-RANS is of type $(1,1,2)$, equivalent to $\mathcal{S}(S_1(R))$ and the two others are of type $(1,2,2)$, translated by 1 with R, which means that the number of their vertices at distance 2 from $O_1(R)$ in is the same as the number of vertices at distance 1 from $O_1(R)$ in R.

This decomposition leads to the following functional equations, with u_j marking vertices at distance j from $O_1(R)$:

$$\begin{aligned} T_2(z,u_1,u_2) &= 1 + zu_1 T_2^2(z,u_1,u_2) F(z,u_2) \\ F(z,u_2) &= 1 + zu_2 G^3(z,u_2), \\ G(z,u_2) &= 1 + zu_2 G(z,u_2) T_1^2(z,u_2). \end{aligned}$$

We go from d to $d+1$ the same way as from 1 to 2 and the preceding equations hold for $d \geq 3$, when considering multivariate generating functions $T(z,u_1,u_2,\ldots,u_d)$, with u_j marking vertices at distance j from $O_1(R)$, and we get the following result.

Proposition 3.1 *Let $r_{n,k_1,\ldots,k_d}$ be the number of RANS of order n with k_j vertices at distance j from O_1. Then $r_{n,k_1,\ldots,k_d}$ is the coefficient of $u_1^{k_1}u_2^{k_2}\ldots u_d^{k_d}z^n$ in the multivariate series $T_d(z,u_1,\ldots,u_d)$, where the series T_d satisfy the recurrence relations:*

$$T_1(z,u_1) = 1 + zu_1 T_1^2(z,u_1) T_0(z) \qquad \textit{with} \qquad T_0(z) = T(z)$$

and for $d \geq 2$,

$$T_d(z,u_1,\ldots,u_d) = 1 + zu_1 T_d^2(z,u_1,\ldots,u_d)\left(1 + zu_2 \frac{1}{(1 - zu_2 T_{d-1}^2(z,u_2,\ldots,u_d))^3}\right).$$

All information concerning distances from vertex O_1 can be retrieved from proposition 3.1:

- The enumerative series for the number of vertices at distance i from O_1, over all RANS, is

$$D_i(z) = \frac{\partial}{\partial u_i} T_i(z,u_1,\ldots,u_i)\Big|_{u_j=1,\forall j} = \sum_n k_i\, r_{n,k_i} z^n.$$

- The total distance from O_1 expresses as

$$\frac{\partial}{\partial u} D(z,u)\Big|_{u=1}, \qquad \text{where } D(z,u) = \sum_{i=1}^{\infty} D_i(z) u^i.$$

The aim of the next paragraph is to evaluate these quantities.

3.2 Distribution analysis

Generating functions counting the number of vertices at distance i from O_1 express as rational functions in z and $T(z)$, and have a singular behaviour similar to $T(z)$: radius of convergence ρ, and singular expansion of the square-root type.

Lemma 3.2 *The sequence of enumerative series for the number of vertices at distance i from O_1 is:*

$$D_1(z) = zT^3(z)/(1-2zT^2(z)),$$
$$D_2(z) = H(z) \times (1+2z^2T^4(z))/(6zT(z)(1-2zT^2(z)))$$
$$\text{and for } i \geq 2 \qquad D_{i+1}(z) = H^{i-1}(z) \times D_2(z),$$

where $H(z) = 6z^2(T(z)-1)T(z)/(1-3z-zT(z)-zT^2(z)+2z^2T^2(z))$[(i)]*. $H(z)$ has a singular expansion $H(z) = 1 - \frac{11}{\sqrt{3}}\sqrt{1-z/\rho+\frac{2}{3}(1-z/\rho)+(1-z/\rho)^{3/2}+O((1-z/\rho)^2)}$, with radius of convergence $\rho = 4/27$.*

Proof: From proposition 3.1, it is easy to compute the expressions of $D_1(z)$ and $D_2(z)$, and show that $D_{i+1}(z) = H(z, T(z)) \times D_i(z)$. The singular expansion comes from expressing z as $(T(z)-1)/T^3(z)$ and plugging in H the singular expansion (1) of $T(z)$. A full expansion of $T(z)$ yields a full expansion for $H(z)$, the first terms of which are given in the lemma. □

Theorem 3.3 *Given R a RANS of order n and v a random internal vertex of R, the distance from v to $O_1(R)$ has a Rayleigh limit distribution:*

$$\Pr(\text{dist}(v, O_1(R)) = x\sqrt{n}) \;=\; c\frac{x}{\sqrt{n}}\, e^{-c^2\frac{x^2}{4}}.$$

Proof: The full singular expansion of $D_i(z)$ can be derived from its expression in terms of H and D_2. Thus the proportion of vertices at distance i from O_1, that is $\frac{1}{nT_n}[z^n]D_i(z)$ can be evaluated:

$$\Pr(\text{dist}(v, O_1(R)) = i) = \frac{1}{nT_n}[z^n]D_i(z) = \frac{1}{nT_n}[z^n]H^{i-2}(z)D_2(z).$$

The result follows from theorem IX.16 (Semi-large powers) of [2]: the singular exponent $1/2$ for $H(z)$ implies a Rayleigh distribution for $k = x\sqrt{n}$. □

The average distance from O_1 is also obtainable by differentiation of $D(z,u) = \sum D_i(z)u^i$. From lemma 3.2 there is a closed form for $D(z,u)$, and differentiation leads (fortunately) to the same series as the one obtained for $\Delta_{①}(z)$, in section 4.1.

This series has a singular expansion around ρ with first term $\frac{3}{44}(1-z/\rho)^{-2}$, so for the mean distance

$$\frac{1}{nT_n}[z^n]\,\frac{\partial}{\partial u}D(z,u)\bigg|_{u=1} = \frac{\sqrt{3\pi n}}{11}\left(1+O(\frac{1}{n})\right).$$

We thus conclude this section with the following proposition:

Proposition 3.4 *In a RANS of order n, the average distance from O_1 is of order $c\sqrt{n}$, with $c = \sqrt{3\pi}/11$.*

(i) Since $\mathbb{Q}[z, T(z)]/\langle T(z) - 1 - zT^3(z)\rangle$ is a $\mathbb{Q}(z)$-vector space with dimension three, all rational functions in z and $T(z)$ that appear in this paper can be expressed in a canonical form. However we didn't use it since it usually hides the combinatorial interpretation of the generating functions under consideration.

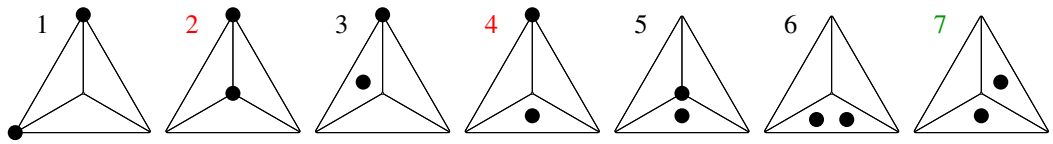

Fig. 3: All possible configurations for pairs of vertices in $R \in \mathcal{R}$. For each pair in $\mathcal{C}(R)$ we can always find a unique sub-RANS of R such that the two vertices are in one of configurations 2, 4 or 7. The other cases reduce to one of these: case 1 reduces to 2 by looking at the RANS containing R, case 3 leads to 2, 3 or 4 by looking in the sub-RANS containing the two vertices, and such is also the case for case 5. Case 6 is amenable to any of 5, 6 or 7 by looking at the sub-RANS containing the two vertices.

4 Total distance between pairs of vertices

In this section, we are interested in computing the total distance of every pair of vertices in a RANS of order n, and will show that the mean value of this quantity is still of order $\sqrt{n}$.

We call $\mathcal{C}(R)$ the set of pairs of vertices (we call pair a set of size two) in $R \in \mathcal{R}$, excluding pairs where both vertices are in $\mathcal{O}(R)$.

The enumerative generating function for the total distance between pairs in $\mathcal{C}(R)$ is

$$G(z) = \sum_{R \in \mathcal{R}} \sum_{(v,w) \in \mathcal{C}(R)} \text{dist}(v, w)\ z^{|R|}.$$

$\mathcal{C}(R)$ splits into two parts

- the pairs (v, w) such that they are both *internal* vertices of the smallest sub-RANS of R that contains both of them, corresponding to case 7 in figure 3.

  We will note them Inter(R) and their contribution to the total distance will be called *interdistance*.

- the others, which can also be defined as the pairs (v, w) such that there exists a sub-RANS S of R with v an outermost vertex of S and w an internal vertex of S, corresponding to cases 2 and 4 in figure 3.

  We will note them Intra(R) and their contribution to the total distance will be called *intradistance*.

Remark: $\mathcal{C}(R)$ has $n(n-1)/2 + 3n$ elements: each pair of internal vertices is counted only once, and the $3n$ term takes into account all pairs made of one internal vertex and one outermost vertex. Among all these pairs, an amount of order $n\sqrt{n}$ belongs to Intra(R) and the rest is in Inter(R). As we will show, the total distance of pairs in Intra(R) is of order n^2 and the total distance of pairs in Inter(R) is of order $n^2\sqrt{n}$. We can thus say that the interdistance gives the dominant term of the total distance in RANS, which is of order $\sqrt{n}$.

We introduce in the following subsection a new generating function which serves as a basis for the computations of all the quantities that are needed. Then we calculate the intradistance followed by the interdistance. Putting everything together gives the following result.

Theorem 4.1 *The mean distance between two vertices in a RANS of order n is asymptotically equivalent to $C\sqrt{n}$, with $C = \sqrt{3\pi}/11$.*

4.1 Cumulated weights generating function

Given $R \in \mathcal{R}$, the distances of inner vertices to $\mathcal{O}(R)$ are denoted by the three following parameters:

$$\Delta_{①}(R) = \sum_{x \in R} d(x, O_1(R)),\ \Delta_{②}(R) = \sum_{x \in R} d(x, \{O_1(R), O_2(R)\}) \text{ and } \Delta_{③}(R) = \sum_{x \in R} d(x, \mathcal{O}(R)).$$

Notice that $\Delta_{①}(R)$ is the sum of all the labels in the $\delta_{①}$-labeling of R, based on RANS of type $(0,1,1)$ —see (3.1). Similarly, $\Delta_{②}(R)$ is the sum of all the labels in the $\delta_{②}$-labeling of R, that starts with a RANS of type $(0,0,1)$; and $\Delta_{③}(R)$ is the sum of all the labels in the $\delta_{③}$-labeling of R, starting with a RANS of type $(0,0,0)$.

In the following generating function, parameter $\Delta_{①}$ is marked by variable $d_{①}$:

$$\Delta(z, d_{①}, d_{②}, d_{③}) = \sum_{R \in \mathcal{R}} d_{①}^{\Delta_{①}(R)} d_{②}^{\Delta_{②}(R)} d_{③}^{\Delta_{③}(R)} z^{|R|} = \sum_{n,i,j,k=0}^{\infty} \alpha_{n,i,j,k}\, d_{①}^{i} d_{②}^{j} d_{③}^{k} z^{n},$$

where $\alpha_{n,i,j,k}$ is the number of RANS of order n (*i.e.* with n internal points), and respective values i, j, k for parameters $\Delta_{①}, \Delta_{②}, \Delta_{③}$. This generating function is called the *cumulated weights generating function* since it expresses the distances according to the three different types of RANS.

Proposition 4.2 *The cumulated weights generating function satisfies the recursive equation*

$$\begin{aligned}\Delta(z, d_{①}, d_{②}, d_{③}) = 1 + z d_{①} d_{②} d_{③} &\times \Delta(z d_{①}, d_{②}, d_{③}, d_{①}) \\ &\times \Delta(z, d_{①}, d_{②} d_{③}, 1) \\ &\times \Delta(z, d_{①} d_{②}, d_{③}, 1).\end{aligned}$$

Proof: Let's follow the recursive definition of RANS R. If R is empty the contribution to the series is 1. Otherwise it has a center, which is at distance 1 from each of the outermost vertices (hence the factor $z d_{①} d_{②} d_{③}$) and the contributions come from the 3 sub-RANS.

Factor $\Delta(z d_{①}, d_{②}, d_{③}, d_{①})$ comes from $S_1(R)$. Suppose $S_1(R)$ has, by itself, a generating series $\Delta(z, d_{①}, d_{②}, d_{③})$, that corresponds to the three different labellings, with types $(0,1,1)$, $(0,0,1)$ and $(0,0,0)$. When it is considered as embedded as the first sub-RANS of R, the top most vertex has label 1 instead of 0, so that the three different labellings now start with types $(1,1,1)$, $(1,0,1)$ and $(1,0,0)$. Thus variable $d_{①}$ transforms into $d_{②}$, variable $d_{②}$ transforms into $d_{③}$, and variable $d_{③}$ transforms into $d_{①}$ with a 1-translation.

Factor $\Delta(z, d_{①}, d_{②} d_{③}, 1)$ comes from $S_2(R)$. Suppose $S_2(R)$ had, by itself, a generating series $\Delta(z, d_{①}, d_{②}, d_{③})$, corresponding to the three different types $(0,1,1)$, $(0,0,1)$ and $(0,0,0)$. When it is considered as embedded as the second sub-RANS of R, $O_2(S_2(R))$ has to have label 1, so that the three different labellings now start with types $(0,1,1)$, $(0,1,1)$ and $(0,1,0)$. Thus variables $d_{②}$ and $d_{③}$ transform into $d_{②}$, variable $d_{①}$ stays $d_{①}$, and nothing transforms to $d_{③}$.

Factor $\Delta(z, d_{①} d_{②}, d_{③}, 1)$ comes from $S_3(R)$, and the proof is equivalent. □

The series of cumulated distances from A is obtained by differentiation

$$\Delta_{①}(z) = \sum_{R \in \mathcal{R}} \Delta_{①}(R) z^{|R|} = \frac{\partial}{\partial d_{①}} \Delta(z, d_{①}, 1, 1)\Big|_{d_{①}=1}$$

and the same holds for the two other cases.

Proposition 4.3 *The distance generating functions $\Delta_{ⓘ}(z)$ have the following expressions:*

$$\begin{aligned}
\Delta_{①}(z) &= zT^3(z)(1-2zT^2(z)+z^2T^4(z)-6z^3T^6(z))/Q(z,T(z))\\
\Delta_{②}(z) &= zT^3(z)(1-3zT^2(z)+4z^2T^4(z)-6z^3T^6(z))/Q(z,T(z))\\
\Delta_{③}(z) &= zT^3(z)(1-3zT^2(z)+2z^2T^4(z))/Q(z,T(z)),\\
&\textit{where } Q(z,T(z)) = (1+2z^2T^4(z))(1-3zT^2(z))^2.
\end{aligned}$$

Each $\Delta_{ⓘ}(z)$ has radius of convergence ρ and a singular expansion of the form:

$$\Delta_{ⓘ}(z) = 3/(44(1-z/\rho)) + O((1-z/\rho)^{-1/2}).$$

Proof: The distance generating functions $\Delta_{①}(z)$ satisfy the system of equations:

$$\begin{cases}
\Delta_{①}(z) &= zT^3(z)+2z\Delta_{①}(z)T^2(z)+zT^2(z)(zT'(z)+\Delta_{①}(z))\\
\Delta_{②}(z) &= zT^3(z)+2z\Delta_{①}(z)T^2(z)+z\Delta_{③}(z)T^2(z)\\
\Delta_{③}(z) &= 3z\Delta_{②}(z)T^2(z)+zT^3(z)
\end{cases}$$

where $T'(z) = T^3(z)/(1-3zT^2(z))$.

The resolution of the system shows that each $\Delta_{①}(z)$ has a dominant term that expresses as $T'^2(z)$ with the same constant factor, thus a pole in $z=\rho$. The singular expansions only differ on their second term. □

4.2 Intradistance

We first consider the pairs (v,w) such that there exists a sub-RANS S of R with v an outermost and w an internal vertex of S. There may be many embedded sub-RANS S and we focus on the smallest one, S_0. In S_0, vertex v is outermost (e.g. O_1) and w is either the center of S_0 or in the sub-RANS opposite to v (e.g. $S_1(S_0)$) (cf. cases 2 and 4 in figure 3).

We will first study the pairs $\text{Intra}_1(R)$, for which $S_0 = R$, and then recursively extend the computation to the rest of the intradistance.

Lemma 4.4 *The generating function for the total distance of pairs in $\text{Intra}_1(R)$, satisfies*

$$\delta(z) = 3T(z) + 3zT^2(z)\Delta_{③}(z) + 3z^2T^2(z)T'(z).$$

Proof: The distance of $\text{Intra}_1(R)$ is made of two categories of distances:

- from the center of R to each of the outermost vertices $\mathcal{O}(R)$,
- from each outermost vertex $O_i(R)$ to all the internal vertices of its opposed sub-RANS, $S_i(R)$.

The distance from an outermost vertex $O_i(R)$ to an internal vertex w of $S_i(R)$ is $1+d(w,\mathcal{O}(S_i(R)))$. Thus, the distance from an outermost vertex $O_i(R)$ to all the internal vertices of $S_i(R)$ is $|S_i(R)| + \Delta_{③}(S_i(R))$. Taking into account all three sub-RANS of R we have

$$\delta(z) = \sum_{R\in\mathcal{R}} \left(3 + \sum_{S\in\mathcal{S}(R)} \left(\Delta_{③}(S) + |S|\right)\right) z^{|R|},$$

thus the expression of the generating function as stated in the lemma. □

Theorem 4.5 *The generating function for intradistances in a RANS is Intra*$(z) = \delta(z)/(1 - 3zT^2(z))$ *and the total distance between pairs of vertices in Intra*(R)*, for* $R \in \mathcal{R}_n$*, is asymptotically* $\frac{3}{11}n^2$.

Proof: The total intradistance is obtained by recursively computing intradistances at any level of the RANS. The effect of this recursion process, akin to recursive descent in subtrees of ternary trees, is to multiply the generating function by $T'(z)/T^3(z)$, that is $1/(1 - 3zT^2(z))$. The dominant term in the singular expansion of Intra(z) thus is $3z\Delta_{③}(z)T'(z)/T(z)$. The total distance in Intra(R) is obtained by evaluating $[z^n]$Intra$(z)/T_n$. □

4.3 Interdistance

We now consider the pairs (v, w) such that they are both *internal* vertices of the smallest sub-RANS S of R that contains both of them.

Since S is minimal by inclusion, v and w are in different sub-RANS of S, which we will call S_v and S_w. The shortest path from v to w can be decomposed in three sub-paths: from v to $\mathcal{O}(S_v)$, from $\mathcal{O}(S_w)$ to w and, if these two sub-paths are disjoint, an one-edge path (on the border of S_v, S_w or S). We will call this last part the *f-edge*. This decomposition is illustrated in figure 4.

We will first compute a lower bound Inter$^-(R)$ of the interdistance by neglecting the f-edges. This lower bound gives a total distance on pairs of Inter(R), with $R \in \mathcal{R}_n$ order $n^2\sqrt{n}$. We can also take an upper bound Inter$^+(R)$ by forcing every path to pass from the center of R and this still gives a contribution of order $n^2\sqrt{n}$ with the same factor. Counting the exact number of f-edges allows us to compute the following terms of the interdistance.

4.3.1 Lower bound and upper bound

As for the intradistance we first compute a lower (resp. upper) bound on interdistances at the topmost level Inter$_1(R)$ (i.e. $S_v, S_w \in \mathcal{S}(R)$) and extend it recursively for the whole RANS.

Lemma 4.6 *The generating function for lower bound (resp. upper bound) of the total distance of pairs in Inter*$_1(R)$*,* $\gamma^-(z)$ *(resp.* $\gamma^+(z)$*), satisfies*

$$\gamma^-(z) = 6z^2T(z)T'(z)\Delta_{③}(z) \qquad \gamma^+(z) = 6z^2T(z)T'(z)\Delta_{①}(z).$$

Proof: At level one, for each sub-RANS, the contribution to the interdistance is the total length of the sub-paths contained in this sub-RANS. Thus for each vertex v, situated in a sub-RANS $S_1(R)$, we will count its distance to $\mathcal{O}(S_1(R))$, multiplied by the number of vertices in $S_2(R)$ and $S_3(R)$. The lower bound is obtained by adding all these values (parameter $\Delta_{③}$), and for the upper bound we consider that

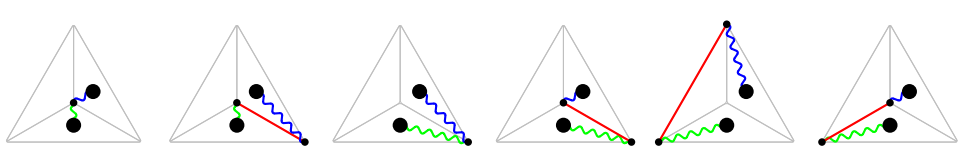

Fig. 4: The different scenarios for paths between pairs of vertices in Inter(R). The three colors correspond to the three sub-paths, the red one being the *f-edge*.

the frontier is reduced to only one of its two points (parameter $\Delta_{①}$). Thus the expression of the generating function $\gamma^-(z)$ (and γ^+ is obtained by replacing $\Delta_{③}$ by $\Delta_{①}$):

$$\gamma^-(z) = 3 \sum_{R \in \mathcal{R}} \Delta_{③}(S_1(R)) \times (|S_2(R)| + |S_3(R)|)\, z^{|T|}. \qquad \square$$

Theorem 4.7 *The generating function for the lower bound (resp. upper bound) of interdistances in a RANS is*

$$\mathit{Inter}^-(z) = \frac{\gamma^-(z)}{1 - 3zT^2(z)} \qquad \mathit{Inter}^+(z) = \frac{\gamma^+(z)}{1 - 3zT^2(z)}$$

and in both cases the total distance between pairs of vertices in Inter(R), for $R \in \mathcal{R}_n$, is asymptotically $Cn^2\sqrt{n}$ with $C = \sqrt{3\pi}/22$.

Proof: The proof is similar to the proof of theorem 4.5. The generating functions Inter$^-(z)$ and Inter$^+(z)$ have both a dominant term in $6\Delta_{①}(z)z^2T'^2(z)/T^2(z)$. $\square$

4.3.2 Exact computation

To count f-edges we consider the shortest way, for a vertex v in S, to reach the outermost vertices of S. Let us denote by $\mathcal{X}(S, v)$ the subset of $\mathcal{O}(S)$ of exit points for v out of S. Let $(v, w) \in$ Inter(R), the shortest path from v to w will contain a f-edge iff $\mathcal{X}(S_v, v) \cap \mathcal{X}(S_w, w) = \emptyset$.

Given a RANS R, we note by $\mathcal{E}_i(R), i \in \{1, 2, 3\}$ the number of vertices inside R with i exit points out of R, divided by $\binom{3}{i}$ (to take symmetries into account). The corresponding generating functions are $E_i(z) = \sum_{R \in \mathcal{R}} |\mathcal{E}_i(R)| z^{|R|}$.

Lemma 4.8 *The generating function for the number of f-edges in the total distance of Inter$_1(R)$ is*

$$\phi(z) = 3zT(z)\left(7E_1^2(z) + 10E_1(z)E_2(z) + 2E_2^2(z) + 2E_2(z)E_3(z)\right).$$

Proof: In order to count the number of pairs in Inter$_1(R)$ for which $\mathcal{X}(S_v, v) \cap \mathcal{X}(S_w, w) = \emptyset$, we need to enumerate all the possible values of $\mathcal{X}(S_v, v)$ and $\mathcal{X}(S_w, w)$. There are seven possibilities for each of them, thus forty-nine combinations from which twenty-one produce a f-edge. The number of times each of the twenty-one combinations arises is $E_i(S_v)E_j(S_w)(|S| - |S_v| - |S_w|)$, if we note $i = |\mathcal{X}(S_v, v)|$ and $j = |\mathcal{X}(S_w, w)|$. $\square$

Calculating $E_i(z)$. We use a multivariate generating function $T_e(z, u_1, u_2, u_3)$ for RANS marked with vertices in $\mathcal{E}_i$. This will be defined by a system of thirty three equations in the same spirit as in section 3.1. The analysis of this system is too long to be included in this abstract, but it leads to rational functions in terms of $T(z)$ and singular expansions around ρ which are respectively equivalent to $\frac{3\sqrt{3}}{22}(1 - z/\rho)^{-1/2}$, $\frac{3\sqrt{3}}{44}(1 - z/\rho)^{-1/2}$, and $\frac{\sqrt{3}}{22}(1 - z/\rho)^{-1/2}$.

Theorem 4.9 *The generating function for the number of f-edges in a RANS is $F(z) = \phi(z)/(1-3zT^2(z))$ and the total number of f-edges in R, for $R \in \mathcal{R}_n$, is asymptotically $\frac{29}{121}n^2$.*

Proof: The proof is similar to theorem 4.5. But in this case, each term of $\phi(z)$ gives a part of the dominant contribution. $\square$

4.4 Conclusion

Summing all contributions, the enumerating generating function for the total distance between pairs of vertices $G(z) = \text{Intra}(z) + \text{Inter}^-(z) + F(z)$ expresses (with $t = T(z)$) as:

$$G(z) = \frac{(180 - 828t + 1428t^2 - 488t^3 - 2029t^4 + 3649t^5 - 2826t^6 + 1086t^7 - 168t^8)t}{(3t^2 - 4t + 2)^2(2t - 3)^4}.$$

We made an exhaustive study of the different parts of $G(z)$. The contribution coming from intradistances happens to be of smaller order (n^2) than the contribution coming from interdistances ($n^2\sqrt{n}$). In the computation of interdistances, we first considered approximations that give a lower and an upper bound with the same dominant term ($\frac{\sqrt{3\pi}}{22}n^2\sqrt{n}$) that is a mean distance in $\frac{\sqrt{3\pi}}{11}\sqrt{n}$. The study of f-edges provides an exact computation of the total distance. With this contribution of f-edges it is possible to express the second term in the asymptotic expression of the total distance. Moreover, relying on full singular expansion of all series under consideration, it is possible to give a full asymptotic expansion of the total distance.

References

[1] A. DARRASSE AND M. SORIA. Degree distribution of random Apollonian network structures and Boltzmann sampling. *2007 Conference on Analysis of Algorithms, AofA 07*, DMTCS Proceedings.

[2] P. FLAJOLET AND R. SEDGEWICK. *Analytic Combinatorics*, web edition, 809+xii pages (available from the authors' web sites). To be published in 2008 by Cambridge University Press.

[3] M.E.J. NEWMAN, A.L. BARABÁSI AND D.J. WATTS. *The structure and dynamics of networks*. Princeton University Press, 2006

[4] T. ZHOU, G. YAN, AND B.-H. WANG. Maximal planar networks with large clustering coefficient and power-law degree distribution journal. *Physical Review E*, 71(4):46141, 2005.

[5] Z. ZHANG, L. CHEN, S. ZHOU, L. FANG, J. GUAN, AND T. ZOU. Exact analytical solution of average path length for Apollonian networks, *arXiv:0706.3491v2 [cond-mat.stat-mech]*, 2007.

[6] M. ALBENQUE AND J.-F. MARCKERT. Some families of increasing planar maps, *arXiv:0712.0593v1 [math.PR]*, 2007.