

# The stable configuration in acyclic preference-based systems

Fabien Mathieu, Gheorghe Postelnicu, Julien Reynier

► **To cite this version:**

Fabien Mathieu, Gheorghe Postelnicu, Julien Reynier. The stable configuration in acyclic preference-based systems. [Research Report] RR-6628, INRIA. 2008, pp.26. inria-00318621

**HAL Id: inria-00318621**

**<https://hal.inria.fr/inria-00318621>**

Submitted on 4 Sep 2008

**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.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

*The stable configuration in  
acyclic preference-based systems*

Fabien Mathieu — Gheorghe Postelnicu — Julien Reynier

N° 6628

September 2008

Thème NUM

*R*apport  
de recherche



## The stable configuration in acyclic preference-based systems

Fabien Mathieu\* , Gheorghe Postelnicu† , Julien Reynier‡

Thème NUM — Systèmes numériques  
Projets GANG

Rapport de recherche n° 6628 — September 2008 — 23 pages

**Abstract:** Acyclic preferences recently appeared as an elegant way to model many distributed systems. An acyclic instance admits a unique stable configuration, which can reveal the performance of the system. In this paper, we give the statistical properties of the stable configuration for three classes of acyclic preferences: node-based preferences, distance-based preferences, and random acyclic systems. Using random overlay graphs, we prove using mean-field and fluid-limit techniques that these systems have an asymptotically continuous independent rank distribution for a proper scaling, and the analytical solution is compared to simulations. These results provide a theoretical ground for validating the performance of bandwidth-based or proximity-based unstructured systems.

**Key-words:** Acyclicity, rank distribution, fluid limit, mean-field, small-worlds, PDE

Supported by CRC MARDI II

Supported by ANR Project Shaman

\* Orange Labs

† Google

‡ Reech AiM Research

# Configuration stable des systèmes à préférences acycliques

**Résumé :** Les systèmes à préférences acycliques sont récemment apparus comme une méthode élégante de modélisation de certains systèmes distribués de type pair-à-pair. Une instance acyclique admet une unique configuration stable, auto-stabilisante, qui donne une bonne indication du comportement du système. Dans ce rapport, nous donnons la distribution statistique de la configuration stable pour trois types de préférences acycliques : les préférences globales (basées sur un ordre total des nœuds), les préférences de distance (le plus proche est préféré), et les préférences acycliques aléatoires. Sous l'hypothèse d'un graphe de compatibilité Erdős-Rényi, nous montrons à l'aide de techniques de limites fluides et de champ moyen l'existence d'une distribution limite continue. La pertinence des résultats est vérifiée à l'aide de simulations.

**Mots-clés :** Systèmes acycliques, distribution, limite fluid, champ moyen, petit-mondes, EDP

## Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Contribution . . . . .	4
1.2	Roadmap . . . . .	4
<b>2</b>	<b>Model and notation</b>	<b>5</b>
<b>3</b>	<b>Acyclic formulas</b>	<b>5</b>
3.1	Generic formula . . . . .	6
3.2	Mean-field approximation . . . . .	6
<b>4</b>	<b>Node-based preferences</b>	<b>6</b>
4.1	Fluid limit . . . . .	7
4.2	Constant degree scaling . . . . .	7
4.3	Convergence theorem . . . . .	7
4.4	Validation . . . . .	8
4.5	Exact resolution . . . . .	8
<b>5</b>	<b>Acyclic and distance-based preferences</b>	<b>9</b>
5.1	Complete rank distribution . . . . .	9
5.1.1	Fluid limit . . . . .	9
5.1.2	Validation . . . . .	10
5.2	Distance distribution . . . . .	10
5.3	Acceptable rank distribution . . . . .	11
<b>6</b>	<b><i>b</i>-matching generalization</b>	<b>13</b>
6.1	Mean Field formulas . . . . .	13
6.2	Fluid limits . . . . .	14
6.3	Discussion . . . . .	14
6.3.1	Stratification trade-off . . . . .	14
6.3.2	Small-World effect in geometric preferences . . . . .	15
<b>7</b>	<b>Conclusion</b>	<b>16</b>
<b>A</b>	<b>Proof of Theorem 1</b>	<b>18</b>
A.1	Uniform convergence . . . . .	18
A.2	PDE . . . . .	19
A.3	Resolution . . . . .	19
<b>B</b>	<b>Exact resolution of the node-based stable configuration</b>	<b>21</b>
B.1	Recursive formula . . . . .	21
B.2	Uniform convergence . . . . .	22
B.3	PDE . . . . .	22

# 1 Introduction

Matching problems with preferences have applications in a variety of real-world situations, including dating agencies, college admissions, roommate attributions, assignment of graduating medical students to their first hospital appointment, or kidney exchanges programs [8, 9, 10, 19, 20].

Recently, matching problems also appeared as an elegant way to model many distributed systems, including ad-hoc and peer-to-peer networks [12, 6, 16, 15, 7]. In distributed systems, the preferences generally come from direct measurements. Those measurements can be node-related (CPU, upload/download bandwidths, storage, battery, uptime), or edge-related (Round-Trip Time, physical/virtual distances, link capacity, co-uptime). In most cases, the resulting preferences are acyclic: there cannot exist a cycle of more than two nodes such that each node prefers its successor to its predecessor. As a consequence, there always exists a unique stable configuration, which is self-stabilizing [6, 1]. This makes things much easier than in other matching problems, where finding, counting and comparing the stable configurations are some of the main issues [8, 21, 17, 20].

Modeling distributed systems with acyclic preferences allows us to predict the effective collaborations that will occur, which, in turn, allows us to infer the performance of a given system. For convenience, the study of an acyclic distributed system is often split into two main problems:

- *How fast is the stabilization process?* Because distributed systems are often highly dynamic, with constant churn and preference alteration, the speed of convergence can be used to determine how far the effective configurations are from the time-evolving stable configuration.
- *What are the properties of the stable configuration?* If the stabilization process is fast enough, the effective and stable configurations will be close. Analyzing the latter can then give valuable information on the former.

In a previous work, Mathieu investigated the first question [16, 15]. He proved that even if the convergence can be prohibitive under an adversary scheduler, it is fast for realistic scenarios. The second question has been answered for specific acyclic preferences: for real-world latency-based preferences, the stable configuration shows, for  $b$ -matching (several mates per nodes allowed), small-world properties (low diameter and high clustering coefficient) [6]; for node-related preferences, the stable configuration tends to pair nodes with similar values [7]: this is the stratification effect, which allows, for instance, to understand upload/download correlations in incentive networks like BitTorrent [4].

## 1.1 Contribution

The studies proposed in [7] and [6] gave only partial, mostly empirical, answers about the link distribution in the stable configuration, and proposed some conjectures. The goal of this paper is to complete and give theoretical proofs on the shape of the stable configuration.

We extend the seminal results that were given in [7] for node-based preferences: for  $b = 1$  (simple matching case), we prove the existence of a limit continuous distribution and solve the corresponding Partial Differential Equation (PDE). Then we apply a similar method for distance-based and random-acyclic preferences, and also give the explicit solution of the corresponding PDE.

Lastly, we extend the results for  $b > 1$  (multiple matchings). In that case, there is no simple expression that gives the exact solution of the PDEs system, but discrete equations are used to observe asymptotical behavior of the distribution. For node-based preferences, the exponential behavior validates the stratification effect (the probability to be matched with a distant peer decreases exponentially with the distance), while the power law obtained for the two other cases indicates that the small world effect observed in [6] for latency is in fact common to all distance-based preferences<sup>1</sup>.

## 1.2 Roadmap

In Section 2 we define the model and notation for preference-based systems. Section 3 gives the generic mean field method used in this paper to solve the simple matching case. The case of node-based preferences is solved in Section 4, then the results are adapted to the distance-based and random-acyclic preferences in Section 5. Section 6 extend the formulas to multiple matchings, and asymptotical properties of the distributions are described. Lastly, Section 7 concludes.

<sup>1</sup>Latencies cannot be considered as real distances, mainly because the triangular inequality is not always verified. However, they form an inframetric, which is no too far from a real metric [13].

## 2 Model and notation

A preference-based system is a set  $V$  of  $N$  nodes, whose possible interactions are described by an acceptance graph  $G$ , a mark matrix  $m$  and a quota vector  $b$ .

The quota vector  $b$  limits the collaborations: a peer  $i$  cannot have more  $b(i)$  simultaneous mates.

The acceptance graph  $G = (V, E)$  is an undirected, non-reflexive graph. It describes allowed matchings: a node  $i$  and a node  $j$  can be mated (we say that  $i$  is acceptable for  $j$ , and *vice versa*) if, and only if (iff)  $\{i, j\} \in E$ . For instance, in peer-to-peer networks, a node cannot be directly connected to all other peers of the system, because of scalability, and peers that are not directly connected cannot be mated. In this paper, we consider Erdős-Rényi graphs  $\mathcal{G}(N, p)$  (each possible edge exists with probability  $p$  independently of the others; hence the expected degree is  $d = p(N - 1)$ ).

The mark matrix  $m$  is used to construct the peers' preferences: given two nodes  $j$  and  $k$  acceptable for  $i$ ,  $i$  ranks  $j$  better than  $k$  iff  $m_{i,j} < m_{i,k}$  (the sign is arbitrary). The following marks are considered in this paper:

**Node-based**  $m(., i)$  is constant (nodes have intrinsic values). These preferences are suited to modeling peer-related performance, like access bandwidth, storage, CPU, uptime. . .

**Geometric** the nodes are associated to  $N$  points picked uniformly at random on a  $n$ -dimensional torus ( $n \geq 1$ ). The marks are the distances between those points. These preferences allow a theoretical analysis of proximity-based performance.

**Meridian latencies** we considered random subsets of  $N$  nodes taken from the 2500 nodes dataset of the Meridian Project [18]. The marks are the (symmetric) latencies between those nodes. We do not perform analysis for those marks, but use them in § 6.3.2 for validating the geometric approach.

**Random acyclic** each edge receives a random uniformly distributed value. The name is justified because all acyclic preferences can be described by marks on the edges (which is equivalent to assume that  $m$  is symmetric). Hence uniformly distributed (symmetric) random marks are a convenient way to perform a uniform sampling of the acyclic preferences [6, 1].

All the considered marks are acyclic, and therefore a  $(G, m, b)$  system admits a unique stable configuration  $C \in E$ , which is self-stabilizing [12, 16]. The neighbors of  $i$  in  $C$  are the stable mates of  $i$ , and the notation  $i \leftrightarrow j$  is used to express that  $i$  and  $j$  are stable mates.

We assume for simplicity that  $m$  is complete and not limited to the edges of  $\mathcal{G}$ . For all considered preferences but random acyclic, the completion is straightforward. For random acyclic preferences, we assume that dummy random values are assigned to non-acceptable edges.

The preferences are denoted like follows: if  $j$  is acceptable for  $i$ ,  $r_i(j)$  denotes the rank of  $j$  in  $i$ 's list (1 being the best).  $r_i$  is called the *acceptable* ranking of  $i$ . If  $i$  has more than  $k$  acceptable neighbors,  $r_i^{-1}(k)$  is the  $k^{\text{th}}$  node in  $i$ 's acceptable ranking. Similarly, for  $j \neq i$ ,  $R_i(j)$  denotes the rank of  $j$  in the complete graph (the acceptability condition is omitted).  $R_i$  is called the *complete* ranking of  $i$ . For  $K < N$ ,  $R_i^{-1}(K)$  is the  $K^{\text{th}}$  node in  $i$ 's complete ranking.

All stable mating probabilities that are discussed in this article are designed by  $D$ . Subscripts and arguments are used to precise the meaning of  $D$  whenever needed. For instance:

- $D_{R_i}(K)$  is the probability that  $i$  has a stable mate with complete rank  $K$ .
- $D_{N,d}(i, j)$  is the probability that  $i \leftrightarrow j$ , knowing there is  $N$  nodes and that the expected degree of the acceptance graph is  $d$ .
- for  $c \leq b(i, .)$ ,  $D_{r_i,c}(k)$  is the probability that the  $c^{\text{th}}$  stable mate of  $i$  has relative rank  $k$ .
- . . .

The complementary cumulative distribution function (CCDF) of  $D$  is denoted  $S$ , and the scaled version of  $D$  and  $S$  are denoted  $\mathcal{D}$  and  $\mathcal{S}$ .

## 3 Acyclic formulas

We first consider the case  $b = 1$  (simple matching) (the results will be extended to multiple matchings in Section 6). We give a generic formula that describes the complete rank of the mate  $C(i)$  of a peer  $i$ .



### 3.1 Generic formula

Let  $D_{R_i}(K)$  be the probability that  $R_i(C(i)) = K$  (the probability that the mate of  $i$ , if any, has rank  $K$ ). The CCDF of  $D$  is  $S_{R_i}(K) := 1 - \sum_{L=1}^{K-1}$ , which is the probability that  $i$ 's mate has a rank greater than  $K$  ( $R_i(C(i)) \geq K$ ) or has no mate (short notation:  $R_i(C(i)) \not\leftarrow K$ ). Following the approach proposed in [7], we first give a generic exact formula that describes  $D_{R_i}$ , then we propose a simplified mean-field approximation.

In order to solve  $D_{R_i}(K)$ , one can observe that  $i$  is mated with its  $K^{\text{th}}$  peer  $j = R_i^{-1}(K)$  iff:

- $\{i, j\}$  is an edge of the acceptance graph; this happens with probability  $p$  as  $G$  is supposed to be a  $\mathcal{G}(N, p)$  graph.
- $i$  is not mated with a node better than  $j$  ( $R_i(C(i)) \not\leftarrow K$ );
- $j$  is not mated with a node better than  $i$  ( $R_j(C(j)) \not\leftarrow R_j(i)$ ).

This leads to the following exact formula:

$$\begin{aligned} D_{R_i}(K) &= p\mathbb{P}(R_i(C(i)) \not\leftarrow K) \times \\ &\quad \times \mathbb{P}(R_j(C(j)) \not\leftarrow R_j(i) | R_i(C(i)) \not\leftarrow K) \\ &= pS_{R_i}(K)\mathbb{P}(R_j(C(j)) \not\leftarrow R_j(i) | R_i(C(i)) \not\leftarrow K) \end{aligned} \quad (1)$$

### 3.2 Mean-field approximation

Solving (1) is difficult to handle, mainly because of possible correlations between  $R_j(C(j)) \not\leftarrow R_j(i)$  and  $R_i(C(i)) \not\leftarrow K$ . The solution is to adopt a mean field assumption:

**Assumption 1** *The events node  $i$  is not with a node better than  $j$  and node  $j$  is not with a node better than  $i$  are independent.*

This assumption has been proposed in [7] to solve (1) in the case of node-based preferences. It is reasonable when  $N$  is large and  $p$  is small. Then (1) can be approximated by

$$D_{R_i}(K) = pS_{R_i}(K)S_{R_j}(R_j(i)). \quad (2)$$

Now, in the next two sections, we propose to solve Equation 2 for specific preferences.

## 4 Node-based preferences

We assume here that the preferences comes from marks on nodes. This is equivalent to assume a total order among the nodes. Therefore we do not need to explicit the mark matrix  $m$ , and we can use an ordered node labeling instead. We arbitrary choose  $1, \dots, N$  as labels, 1 been the best (if 1 is ranked first for all nodes that accept 1, and so on...).

Because the nodes' label express their complete ranks, we can directly consider  $D(i, j)$ , the probability that node  $i$  is mated with node  $j$ . Node  $j$  has rank  $j$  for  $i$  if  $j < i$ , and  $j - 1$  if  $j > i$ , because a rank does not rank itself. This gives the relation between  $D$  and  $D_R$ :

$$D(i, j) = \begin{cases} D_{R_i}(j) & \text{if } j < i, \\ 0 & \text{if } j = i \text{ (mating is not reflexive),} \\ D_{R_i}(j - 1) & \text{if } j > i. \end{cases} \quad (3)$$

Using the CCDF  $S(i, j) := 1 - \sum_{k=1}^{j-1} D(i, k)$ , we get the node-based version of Equation 2:

$$D(i, j) = \begin{cases} 0 & \text{if } i = j, \\ pS(i, j)S(j, i) & \text{otherwise.} \end{cases} \quad (4)$$

This equation, which was originally proposed in [7], which also show that it gives a very good approximation of empirical distribution. It can be numerically solved by using a double iteration.

## 4.1 Fluid limit

Our main contribution for node-based preferences is to prove that, under a constant degree scaling,  $D$  admits a fluid limit. This limit gives a complete description of  $D$  that can be applied to all values of  $N$  and  $p$ , while Equation (4) needs to be solved for each set of parameters.

## 4.2 Constant degree scaling

In order to compare the distributions for arbitrary values of  $N$ , we need a scaled version of  $D$ , where a peer  $i$  is represented by a scaled ranking  $0 \leq \alpha < 1$ . In details, we associate to each  $i$  the number  $\alpha(i) = \frac{i-1}{N}$ , and to each real number  $\alpha$  the node  $i(\alpha) = \lfloor N\alpha \rfloor + 1$ . The scaled version of  $D$ , denoted  $\mathcal{D}$ , is then defined by

$$\mathcal{D}_N(\alpha, \beta) = ND(\lfloor N\alpha \rfloor + 1, \lfloor N\beta \rfloor + 1).$$

$\mathcal{D}_N$  is a piecewise constant function. Its set of function values is the set of the  $(ND(i, j))$  values. The factor  $N$  in its definition allows to express  $D(i, j)$  as an integral of  $\mathcal{D}$ :

$$D(i, j) = \int_{\frac{i-1}{N}}^{\frac{j}{N}} \mathcal{D}_N\left(\frac{i-1}{N}, x\right) dx = \int_{\frac{i-1}{N}}^{\frac{j}{N}} \mathcal{D}_N\left(x, \frac{j-1}{N}\right) dx$$

The scaling of the CCDF is defined by

$$\mathcal{S}_N(\alpha, \beta) = 1 - \int_0^\beta \mathcal{D}_N(\alpha, x) dx, \quad (5)$$

and the relation between  $S$  and  $\mathcal{S}$  is

$$S(i, j) = \mathcal{S}\left(\frac{i-1}{N}, \frac{j-1}{N}\right). \quad (6)$$

## 4.3 Convergence theorem

We now want to show the existence of a continuous limit for  $\mathcal{D}$ . The problem is the existence of a discontinuity for  $\alpha \approx \beta$ , because  $D(i, i) = 0$ . However, this discontinuity is just a reminder of the fact that a node cannot mate with itself, so we propose to make  $\mathcal{D}$  more ‘‘continuous’’ by introducing

$$\tilde{\mathcal{D}}(\alpha, \beta) = \begin{cases} \mathcal{D}(\alpha, \beta) & \text{if } \lfloor N\alpha \rfloor \neq \lfloor N\beta \rfloor, \\ Np(S(\lfloor N\alpha \rfloor + 1, \lfloor N\alpha \rfloor + 1))^2 & \text{otherwise.} \end{cases}$$

The fluid limit of  $\tilde{\mathcal{D}}$  is then given by the following theorem:

**Theorem 1** *Let  $d > 0$  be a constant. If  $N \rightarrow \infty$  with  $p = \frac{d}{N}$ , the function  $\tilde{\mathcal{D}}_{N,d}$  uniformly converge towards*

$$\mathcal{D}_\infty(\alpha, \beta) = \frac{de^{d(|\beta-\alpha|)}}{(1 - e^{-d\min(\alpha,\beta)} + e^{d|\beta-\alpha|})^2}. \quad (7)$$

This result indicates that asymptotically, the average degree in the acceptance graph completely defines the mating distribution. The consequence is that we can explicitly describe the so-called *stratification* effect [7]: the mating distribution is exponentially decreasing with  $|\beta - \alpha|$ , with intensity  $d$ . In other words, a peer with scaled rank  $\alpha$  tends to mate with a mate of same scaled rank, with a standard deviation of the same order than  $\frac{1}{d}$ .

The proof of Theorem 1 is given in Appendix A. Note, that the existence of a fluid limit was proposed as a conjecture in [7], and proved for  $\alpha = 0$  (but the expression of the fluid limit in the general case was not provided).

Theorem 1 gives two corollaries:

- using the CCDF of  $\mathcal{D}_\infty$ , the probability that a node of scaled rank  $\alpha$  has no mate is  $\frac{1}{1+e^{-d\alpha}(e^{-d}-1)}$ ;
- for  $i \neq j$  (discrete case), a good approximation for  $D(i, j)$  is

$$D(i, j) \approx \frac{pe^{p(|j-i|)}}{(1 - e^{-p\min(i,j)} + e^{p|j-i|})^2}. \quad (8)$$

#### 4.4 Validation

We compared our fluid limit approximation, given by (8), to the mean-field values given by (4), which are known to be accurate ([7]).

$N$  was set to 50 or 2000, and  $d$  to 5 or 30. Because  $D$  is 2-dimensional, we arbitrary set the scaled rank  $\alpha$  to 0.1 or 0.9 (but the convergence validation holds for any  $\alpha$ ). The results are shown in Figure 1.

We observe a gap for  $j = \lfloor N\alpha \rfloor + 1$ , because the mean field formula sets  $D$  to 0 whereas the fluid limit uses a continuous extension.

Besides this gap,  $N = 50$  (Figures 1a and 1b) shows some difference between the mean field and the fluid limit. The error is especially noticeable for  $d = 30$  (1b). However For  $N = 2000$  (Figures 1c and 1d), there is practically no error.

These results are consistent with (29) (in the Appendix A), which shows that the convergence is  $O(\frac{d^2}{N}e^{8d})$ .

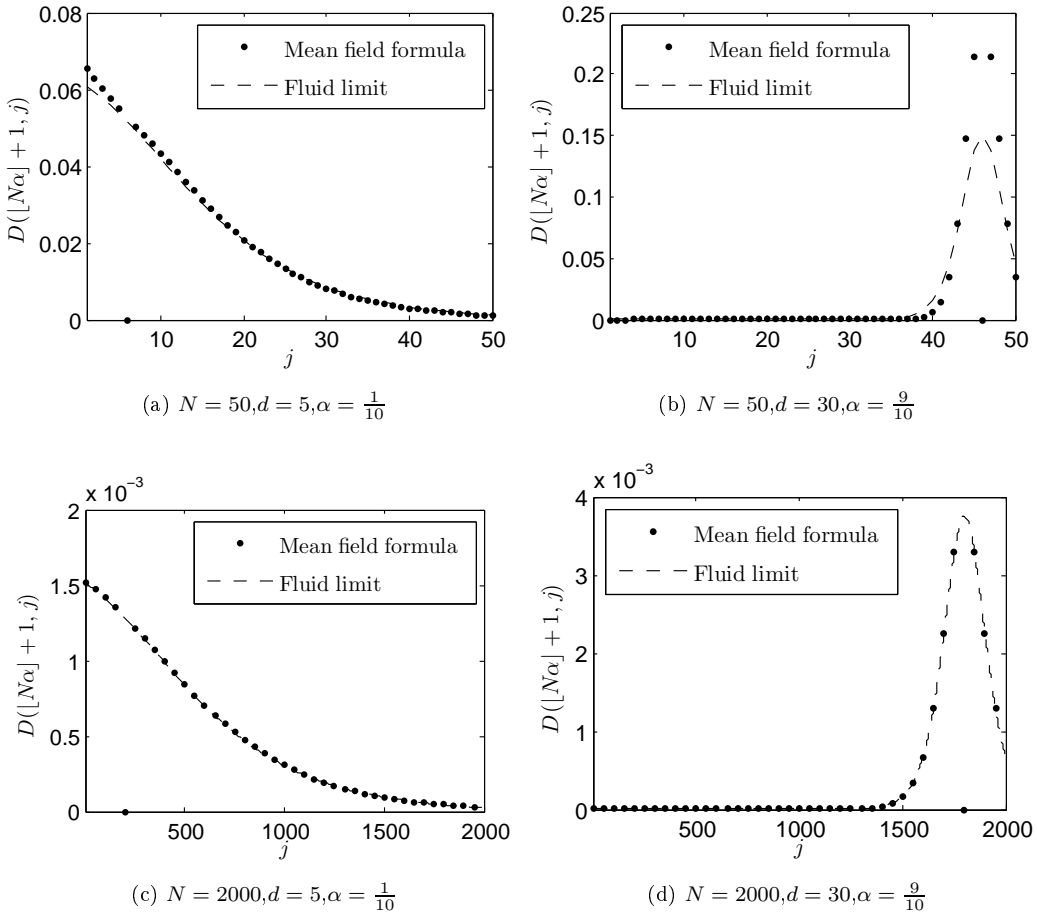


Figure 1: Validation of the fluid limit for node-based preferences.

#### 4.5 Exact resolution

For the record, if  $b = 1$ , there exists an exact recursive formula for the node-based stable configuration. This formula is

$$\begin{aligned}
 D(i, j) = & (1 - S(1, i)) D(i - 2, j - 2) \\
 & + (S(1, i + 1) - S(1, j)) D(i - 1, j - 2) \quad \text{for } i < j, \\
 & + S(1, j + 1) D(i - 1, j - 1)
 \end{aligned} \tag{9}$$

with the border conditions  $D(1, k) = p(1 - p)^{k-2}$  ( $k \geq 2$ ),  $D(i, i) = 0$ .

This equation also admits a fluid limit, which happens to be the same than the fluid limit of the mean field formula. This result appears as a strong validation of the mean field approach: although the mean field formula is not exact (its results differ from the exact formula), its fluid limit is exact.

One could wonder why using a mean field formula if a usable exact formula exists. The issue with the exact formula is that it relies on a “trick”: if you remove node 1 and its mate from the system, the remaining nodes still form a preference-based system with same parameters except there is two less nodes. However, this trick cannot be generalized for other preferences or for  $b > 1$ . This is why we focus on the mean field formulas.

A complete proof of the exact recursive formula, including its PDE counterpart and resolution, can be found in Appendix B.

## 5 Acyclic and distance-based preferences

We now consider geometric and random acyclic preferences. Following the approach used for node-based preferences, we first focus on the complete rank distribution. Use mean field assumptions, we propose a recursive formula for  $D$ , then we solve the fluid limit. The results are then extended to the distance and acceptable rankings distributions.

### 5.1 Complete rank distribution

Assumption 1 is not enough for solving (2) in the case of geometric or random acyclic preferences. Therefore, we propose this additional assumption:

**Assumption 2** *For geometric and random acyclic preferences, the following approximations hold:*

- $D_{R_i}(K)$  is independent of  $i$  (and therefore denoted  $D_R(K)$ );
- the complete ranking is symmetric:  $R_i(j) = R_j(i)$ .

The first approximation just states that in average, all nodes have the same mate distribution, while the second one tells that  $R_i(j)$  is a good approximation of  $R_j(i)$ . These approximations were motivated by the uniform distributions used for shaping the preferences. In particular, they do not apply for node-based preferences, where the mate distribution is strongly affected by a node’s mark. Under these assumptions, we get

$$D_R(K) = pS_R^2(K), \text{ with } S_R(K) = 1 - \sum_{L=1}^{K-1} D_R(L). \quad (10)$$

This equation gives an immediate recursion for  $S_R$ :

$$S_R(K) = \begin{cases} 1 & \text{if } K = 1, \\ S_R(K-1) - pS_R^2(K-1) & \text{otherwise.} \end{cases} \quad (11)$$

In return,  $D_R$  is directly given by  $D_R(K) = S_R(K) - S_R(K+1)$ .

#### 5.1.1 Fluid limit

We now give the fluid limit of  $D_R$ . The scaled version of  $D_R$  is defined like for node-based preferences, except that there now only one parameter. For  $0 \leq \alpha < 1$ , we define  $\mathcal{D}_R(\alpha) := (N-1)D_R(\lfloor (N-1)\alpha \rfloor + 1)$ . The scaling factor is now  $N-1$  because it is the upper bound for  $K$  (while  $N$  was the upper bound for  $i, j$  in §4).  $D_R$  can be expressed as an integral of  $\mathcal{D}_R$ :  $D_R(K) = \int_{\frac{K-1}{N-1}}^{\frac{K}{N-1}} \mathcal{D}_R(x) dx$ . The scaled CCDF,  $\mathcal{S}_R$ , is then naturally defined as:

$$\mathcal{S}_R(\alpha) = 1 - \int_0^\alpha \mathcal{D}_R(x) dx.$$

**Theorem 2** *We assume that  $d = p(N-1)$  is a positive constant. As  $N \rightarrow \infty$ ,  $\mathcal{S}_R$  uniformly converges towards*

$$\mathcal{S}_\infty(\alpha) = \frac{1}{d\alpha + 1}. \quad (12)$$

In particular, the probability that a node has no mate in the stable configuration is  $\mathcal{S}_R(1) = \frac{1}{d+1}$ , and a good approximation for  $\mathcal{S}_R(K)$  is

$$\mathcal{S}_R(K) = \frac{1}{p(K-1)+1}. \quad (13)$$

**Sketch of proof:** The proof is a simpler version of the proof of Theorem 1 (cf Appendix A). First we prove that the  $\mathcal{D}_R$  functions are uniformly Cauchy (but in this case there is only one variable and there is no need for a continuous extension). This proves the uniform convergence towards  $\mathcal{S}_\infty$ . Then we deduce from (11) a differential equation verified by  $\mathcal{S}_\infty$ :

$$-\dot{\mathcal{S}}_\infty(\alpha) = d\mathcal{S}_R^2(\alpha), \quad (14)$$

with the boundary condition  $\mathcal{S}_R(0) = 1$ . The resolution of (14) gives (12), which completes the proof.  $\square$

### 5.1.2 Validation

Contrary to the case of node-based preferences, the mean field formula (10) has not been validated in a previous work, so we could not compare the fluid limit with it, and used simulations<sup>2</sup>. We considered random acyclic instances, and geometric preferences in a 1-dimensional torus and in a 6-dimensional torus.  $N$  was set to 50 or 2000. We used 3 values of  $p$ : 1,  $\frac{1}{10}$  and  $\frac{1}{100}$ . For each set of parameters, the empirical distribution was calculated over 100 instances. The results are shown in Figure 2.

For  $p = 1$  (Figures 2a and 2b), the mean-field assumptions hardly hold. As a consequence, the curves depend of the type of preferences, and the fluid limit is not accurate. This is especially visible if  $K$  is close to the boundaries (that is 1 or  $N$ ). In particular, the non-mate probability is clearly over-estimated. However, the fluid limit manages to give the  $O(\frac{1}{K})$  behavior that is common to all considered preferences. From that point of view, the fluid limit performs better than the recursive equation (11), which gives  $\mathcal{S}_R(K) = \delta_K^1$  for  $p = 1$ .

For  $p = \frac{1}{10}$  (Figures 2c and 2d), the curves are nearly indistinguishable. We verify that all types of preferences (acyclic or geometric) tend to have the same behavior and that Theorem 2 gives precise approximations.

For  $p = \frac{1}{100}$  (Figures 2e and 2f), the curves are indistinguishable.

We conclude that fluid-limit based on the mean-field formula is very effective for computing the complete ranking distribution, even if  $N$  is not very large and  $p$  is not very small.

## 5.2 Distance distribution

For geometric preferences, the actual distance between a node and its mate may be a more valuable performance indicator than the ranking. We call  $S_X(x)$  the probability that the distance between a node  $i$  and its mate  $C(i)$  is not less than  $x$  (in other words, the distance is greater than  $x$  or  $i$  is unmated). Under the fluid limit, we get a good estimate of  $S_X$ :

$$S_X(x) = \frac{1}{dB_n(x)+1}, \quad (15)$$

where  $B_n$  is the size of a ball of radius  $x$  in the  $n$ -torus.

**Proof:** In the fluid limit, a ball of radius  $x$  contains  $NB_n(x)$  nodes, because it occupies a ratio  $B_n(x)$  of the torus. Therefore the farthest node in a  $x$ -ball centered at a node  $i$  should have a complete rank  $NB_n(x)$  for  $i$ , while being at a distance  $x$  from  $i$ . We deduce that  $S_X(x) = \mathcal{S}_R(NB_n(x))$ . Equation (13) concludes.  $\square$

The value of  $B_n(x)$  depends on  $n$  and on the norm used. If we consider the maximum norm, then  $B_n(x) = \min((2x)^n, 1)$ . For other norms, the formula may be more complicated because the ball may partially overlap itself in the torus. Note, that if we choose  $\mathbb{R}^n$  (with uniform point distribution) instead of the  $n$ -torus,  $B_n(x)$  is just the size of a ball of radius  $x$ .

Figure 3 shows  $S_X$  for  $n = 1$  and  $n = 3$ , with the taxicab norm. With this norm, we have  $B_1(x) = \min(2x, 1)$  and

$$B_3(x) = \begin{cases} \frac{4}{3}x^3 & \text{if } 0 \leq x \leq \frac{1}{2}, \\ \frac{4}{3}x^3 - 4(x - \frac{1}{2})^3 & \text{if } \frac{1}{2} \leq x \leq 1, \\ 1 - \frac{4}{3}(\frac{3}{2} - x)^3 & \text{if } 1 \leq x \leq \frac{3}{2}, \\ 1 & \text{if } x \geq \frac{3}{2}. \end{cases}$$

We used  $N = 2000$  and  $p = \frac{1}{100}$ , and the fluid limit and empirical distribution of  $S_X$  were indistinguishable.

<sup>2</sup>Actually, we did validate the mean field formula, but our results are to be published.

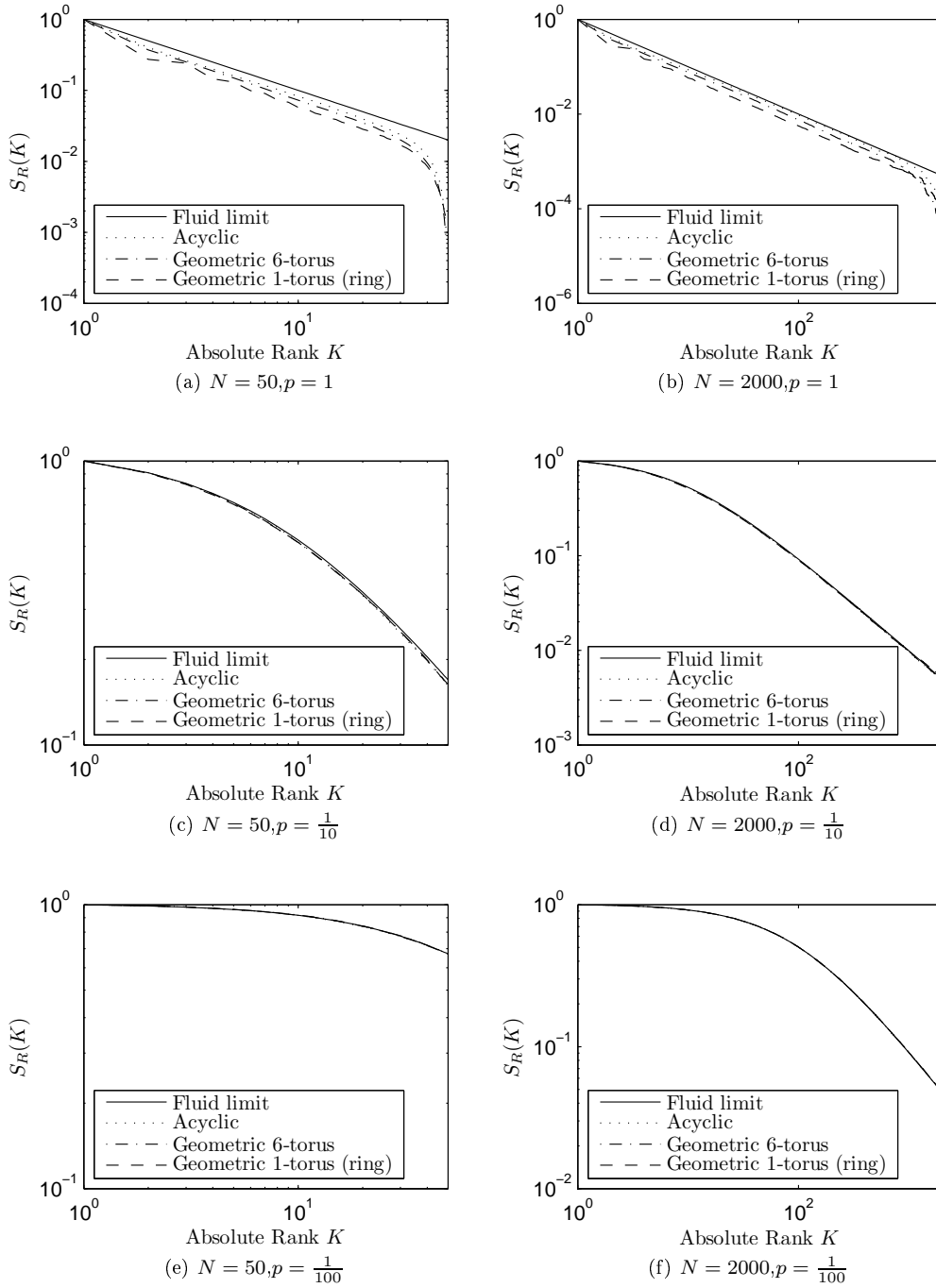


Figure 2: Empirical validation of the fluid limit.

### 5.3 Acceptable rank distribution

Now we want to investigate the probability that the mate of a node has an acceptable rank  $k$ . We call  $D_r(k)$  this probability. Like for the other distributions, we introduce the CCDF  $S_r(k) := 1 - \sum_{i=1}^{k-1} D_r(k)$ .

Following the complete ranking method, we consider the conditions for a node  $i$  to be mated with its  $k^{\text{th}}$  best neighbor  $j = r_i^{-1}(k)$ :

- $i$  must have  $k$  neighbors or more,
- it must not be mated with someone better than  $j$ ,

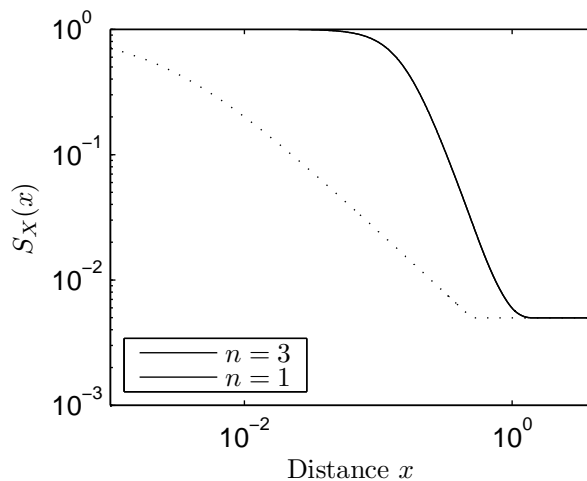


Figure 3: Distance distribution ( $N = 2000, p = \frac{1}{100}$ ).

- $j$  must not be mated with someone better than  $i$ .

With the acceptance ranking, there is intrinsic correlations between these events that complicates things. Despite of that, assuming that these events are independent allows us to give a first, non-accurate, recursive formula:

$$D_r(k) = S_r(k) \frac{1 - I_{1-p}(n - k + 1, k)}{k + 1}, \quad (16)$$

where  $I_x$  is the regularized incomplete beta function.

**Proof:**  $i$  has  $k$  neighbors or more with probability  $1 - I_{1-p}(n - k + 1, k)$ . The probability that  $i$  is not with better than  $j$  is  $S_r(k)$ . For the reciprocal, we can use  $K = \frac{k}{p}$  as a (very rough) approximation of the complete rank; then Equation (13) gives the probability  $\frac{1}{k+1}$ . Formula (16) follows.  $\square$

The results are shown in Figure 4. One can observe that Equation (16) is not accurate for  $D_r(1)$ , which provokes a gap between the empirical distribution and the formula.

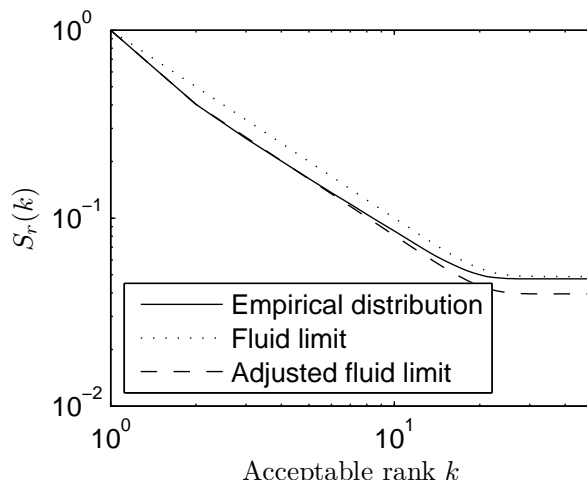


Figure 4: Acceptable ranking CCDF ( $N = 2000, p = \frac{1}{100}$ ).

In an attempt to adjust the formula, we propose a more accurate estimation of  $D_r(1)$ : under the normalized fluid limit, the scaled rank of the first neighbor  $j$  of a given peer  $i$  follows the distribution  $de^{-d\alpha}$ .  $j$  and  $i$  are mate if  $j$  is mated with someone with a scaled rank greater or equal to  $\alpha$ , which happens with probability  $\frac{1}{d\alpha+1}$ .

Thus we have

$$\begin{aligned}
 D_r(1) &= \int_0^\infty \frac{de^{-d\alpha}}{d\alpha+1} d\alpha \\
 &= \int_1^\infty \frac{e^{-t+1}}{t} dt \\
 &= eE_1(1) \approx 0.596
 \end{aligned} \tag{17}$$

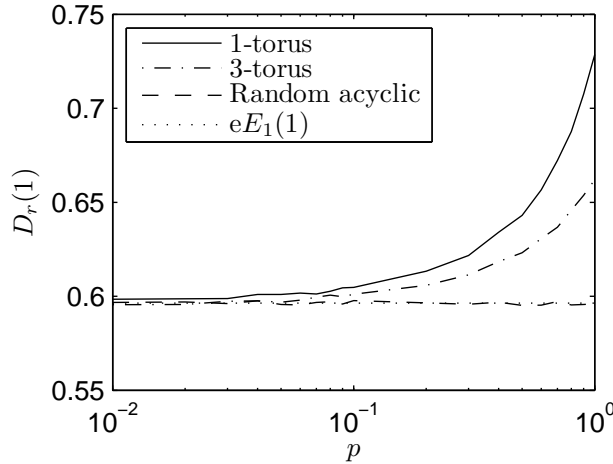


Figure 5:  $D_r(1)$  as a function of  $p$  ( $N = 2000$ ).

The accuracy of  $D_r(1) = eE_1(1)$  ( $E_1$  denotes the exponential integral) is verified in Figure 5. If we use this value for adjusting the fluid limit, we get a better estimation of  $S_r$  for small values of  $k$  (cf Figure 4). However, this adjustment introduces a gap for larger values of  $k$ . In a further version of this paper, we will aim at unifying these two estimates, which will require a better understanding of the correlations that occur when considering the acceptable rank.

## 6 $b$ -matching generalization

We now extend our results to the case of multiple matchings. For simplicity, we consider here that the quota vector  $b$  is a scalar, i.e. that all nodes share the same number of authorized collaborations. For distance and acyclic preferences, we focus on the complete rank, although distance and acceptable ranking could be derived using the same techniques than for  $b = 1$ .

### 6.1 Mean Field formulas

A peer can now have up to  $b$  mates. For  $1 \leq c \leq b$ ,  $D_c$  denotes the distribution of the complete ranking of the  $c^{th}$  best mate, and  $S_c$  denotes the corresponding CCDF. Like we did for  $b = 1$ , we can give the conditions for a node  $j = R_i^{-1}(K)$  to be the  $c^{th}$  mate of a node  $i$ :

- $\{i, j\}$  is an acceptable edge,
- the  $(c - 1)^{th}$  mate of  $i$  (if  $c > 1$ ) is better than  $j$ , but the  $c^{th}$  (if any) is not,
- the  $b^{th}$  mate of  $j$  (if any) is not better than  $i$ .

By extending Assumption 1, we obtain a generic mean field formula for multiple matchings:

$$D_{R_i,c}(K) = \begin{cases} pS_{R_i,1}(K)S_{R_j,b}(R_j(i)) & \text{if } c = 1, \text{ otherwise} \\ p(S_{R_i,c}(K) - S_{R_i,c-1}(K))S_{R_j,b}(R_j(i)). \end{cases} \tag{18}$$

Like for the simple matching case, this formula can be adapted to specific preferences.



We first consider node-based preferences.  $D_c(i, j)$  being the probability that the  $c^{\text{th}}$  mate of  $i$  is  $j$ , we have the following system, which can be solved by a double iteration on  $i$  and  $j$  (cf [7])

$$D_c(i, j) = \begin{cases} 0 & \text{if } i = j, \\ pS_1(i, j)S_b(j, i) & \text{if } i \neq j, c = 1, \\ p(S_c(i, j) - S_{c-1}(i, j))S_b(j, i) & \text{if } i \neq j, c > 1. \end{cases} \quad (19)$$

Then, for acyclic and distance-based preferences, we also extend Assumption 2 (homogeneity of the distributions and symmetry of the complete ranking). This gives the following system:

$$D_{R,c}(K) = \begin{cases} pS_{R,1}(K)S_{R,b}(K) & \text{if } c = 1, \\ p(S_{R,c}(K) - S_{R,c-1}(K))S_{R,b}(K) & \text{if } c > 1. \end{cases} \quad (20)$$

Using  $S_{R,c}(1) = 1$  and  $D_{R,c}(K) = S_{R,c}(K) - S_{R,c}(K+1)$ , Equation (20) immediately gives an iterative computation of  $S_{R,c}$ .

Figure 6 shows  $S_c(i, j)$  (node-based) and  $S_{R,c}$  (acyclic/geometric) as obtained by (19) and (20). The parameters are  $b \in \{2, 3, 4\}$ ,  $N = 2000$ ,  $p = \frac{1}{100}$ , and  $i = 1001$  (for  $S_c(i, j)$ ). We verified for each set of parameters that the curves coincide with the empirical distribution.  $S$  and  $S_R$  (CCDF for  $b = 1$ ) are also plotted for serving as a landmarks. We see that the curves have a behavior that is similar than for the simple matching case: for node-based preferences, it seems that the distribution  $D_c(i, \cdot)$  are still exponentially decreasing, even if seems that there is now offsets between the distribution peaks and  $i$ . For acyclic and geometric preferences, we still observe a kind of power law behavior.

## 6.2 Fluid limits

Fluid limits also exist for  $b > 1$ . We will not present the proofs in this paper, because they are essentially the same that the uniformly Cauchy proofs for the simple matching fluid limits, only more complex to write because of the multiple distribution involved. Therefore we just give the equations verified by the limits.

For node-based preferences, the scaled limit  $\mathcal{S}$  of the CCDF verifies:

$$\partial_y \mathcal{S}_c(\alpha, \beta) = \begin{cases} -d\mathcal{S}_1(\alpha, \beta)\mathcal{S}_b(\beta, \alpha) & \text{for } c = 1, \text{ otherwise} \\ -d(\mathcal{S}_c(\alpha, \beta) - \mathcal{S}_{c-1}(\alpha, \beta))\mathcal{S}_b(\beta, \alpha), \end{cases} \quad (21)$$

with border conditions  $\mathcal{S}_c(\alpha, 0) = 1$ .

Similarly, for acyclic and distance-based preferences, the scaled limit  $\mathcal{S}_R$  of the CCDF verifies

$$\dot{\mathcal{S}}_{R,c} = \begin{cases} -d\mathcal{S}_{R,1}\mathcal{S}_{R,b} & \text{if } c = 1, \\ -d(\mathcal{S}_{R,c} - \mathcal{S}_{R,c-1})\mathcal{S}_{R,b} & \text{if } c > 1, \end{cases} \quad (22)$$

with the boundary condition  $\mathcal{S}_{R,c}(0) = 1$ .

There is no simple explicit solution for Equations (21) and (22). However, (19) and (20) can still be used as difference equations to approximate a numerical solution. The reason for which we give these limits is that we think that they can give us valuable information about the asymptotical behavior of the distribution (exponentially decreasing or power law), even if this work is still to be done.

## 6.3 Discussion

### 6.3.1 Stratification trade-off

As we have seen, for node-based preferences, the mates of a given peer  $i$  have, in average, the same rank than  $i$ . This is the stratification effect ([7]), which guarantees a some fairness in the stable configuration: the expected gain of a node tends to be the value offered by this node, measured in term of ranking. However, we also observed that the exponential decreases of the  $\mathcal{D}_c(i, \cdot)$  functions provokes a standard deviation of the same order that  $\frac{1}{d}$ , where  $d$  is the average degree in the acceptance graph. This gives the following stratification trade-off:

- if  $d$  is too small, the standard deviation is high. In particular, if the mark matrix is non uniformly distributed, there can be a big difference between the expected gain and gift, measured with the marks. This issue has been highlighted in [7] for explaining a possible workaround of BitTorrent's Tit-for-Tat policy;

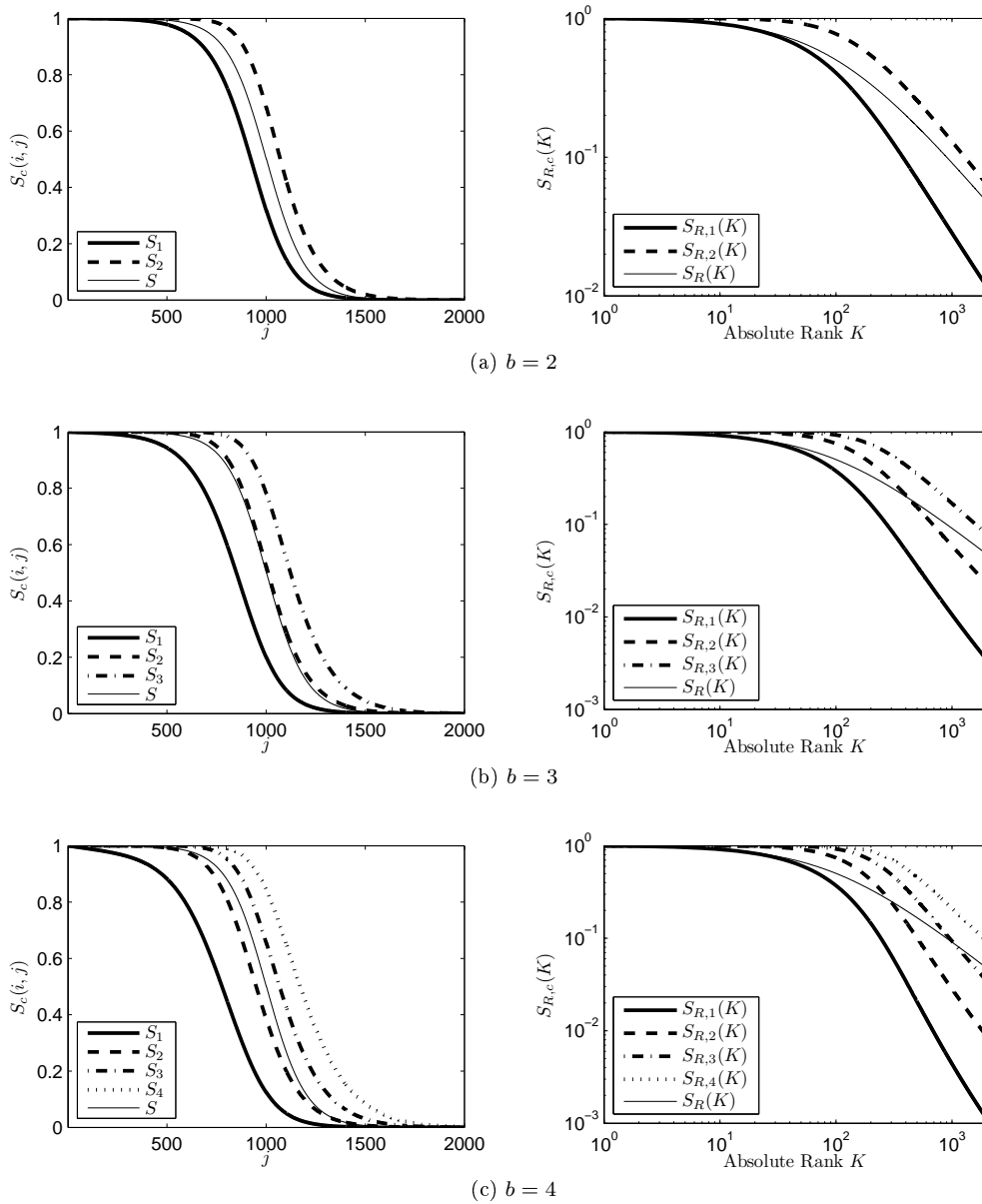


Figure 6: Complete rank CCDFs for  $b > 1$ . Node-based (resp. acyclic/geometric) distributions are on the left (resp. right) side.  $N = 2000$ ,  $p = \frac{1}{100}$ , and  $i = 1001$  (for  $S_c(i, j)$ )

- on the other hand, a high  $d$  will enforce the fairness. However, the size of the acceptance graph degree has a cost for the nodes (memory usage, overlay management, ...). Also, the absence of long-range mates makes the diameter of the stable configuration high, which can be problematic if messages are to be spread using stable edges.

Note, that there is a similar trade off for  $b$ , which is the maximal degree in the stable configuration. This suggests that most node-based preference systems (this includes the systems based on the sharing of an access bandwidth, a storage or CPU capacity, an expected uptime, ...) should admit an optimal pair  $(d, b)$  with respect to the stable collaborations properties, whose values depend on the weight put on the effects presented above.

### 6.3.2 Small-World effect in geometric preferences

A small world is a sparse graph with a low average shortest path length (ASPL) and a high clustering coefficient. In details:

Type of preferences	ASPL	Clustering Coefficient
1-torus	7.4	0.055
2-torus	6.7	0.043
Meridian	6.1	0.031
3-torus	5.9	0.033
4-torus	5.1	0.027
Random Acyclic	5	0.0043

Table 1: ASPL and clustering ( $N = 2000, p = \frac{1}{10}, b = 10$ ).

- *sparse graph* means that the average degree is  $O(1)$  or  $O(\log n)$ ,
- *low ASPL* means  $O(\log(n))$ ,
- *high clustering coefficient* means that two nodes sharing an edge are likely to have a common neighbor. The clustering coefficient is a probability, that must be compared to the clustering coefficient of a random graph with same number of nodes and edges.

In [11], Kleinberg proved that a  $n$ -dimensional grid can be turned into a small world by adding long-range edges that follow a  $\Omega(\frac{1}{x^n})$  distribution.

For multiple matchings, the stable configuration in geometric preferences is likely to have a high clustering coefficient, because most of the stable edges link close nodes. Moreover, the power-law rank distribution tells that long-range edges exist. So the stable configuration is likely to be eligible as a small-world.

In Table 1, we give the ASPL and clustering coefficient for some preferences, using the parameters  $N = 2000, p = \frac{1}{10}, b = 10$ . The reference clustering is here  $\frac{b}{N-1} \approx 0.005$ . We verify that for the  $n$ -tori, the stable configurations are small-worlds. On the other hand, like previously observed in [6], the stable configurations of random acyclic preferences are not small-worlds, because of their clustering coefficient (they behave like an incomplete  $b$ -regular graph).

We also calculated the ASPL and clustering obtained by using the Meridian Project's real-world latencies, which are known to produce small-worlds configuration [6]. One can observe that the results are very close to the one obtained with the tori. Interestingly, the closest results are those from the 3-torus, suggesting that somehow, 3 may be seen as sort of dimension for the latency space. Considering the recent eager for estimating the Internet dimension (see for instance [2]), this unexpected result is appealing: it suggests that the stable configuration, which is only defined by how the nodes rank each other (latencies are used for sorting the nodes, but the actual values are never involved in the construction), could reveal valuable insight about the topology behind a set of distances.

## 7 Conclusion

We gave a statistical description of the stable configurations obtained from node-based preferences, distance-based preferences, and from random acyclic preferences. Starting from a generic formula for the rank distribution, we introduced mean-field and fluid limit techniques in order to give explicit formulas. All our results were validated by means of simulations. An interesting consequence of our results is that for distance-based preferences, the stable configurations behave similarly to Kleinberg's grids, and are small-world graphs.

## References

- [1] D. J. Abraham, A. Levavi, D. Manlove, and G. O'Malley. The stable roommates problem with globally-ranked pairs. In *WINE*, volume 4858 of *Lecture Notes in Computer Science*, pages 431–444. Springer, 2007.
- [2] B. Abrahao and R. Kleinberg. On the internet delay space dimensionality. In *Proceedings of the 2008 Internet Measurement*, 2008.
- [3] R. Bellman. The stability of solutions of linear differential equations. *Duke Math. J.*, 10:643–647, 1943.
- [4] B. Cohen. Incentives build robustness in bittorrent. In *P2PECON*, 2003.
- [5] L. C. Evans. *Partial Differential Equations*. American Mathematical Society, 1998.
- [6] A.-T. Gai, D. Lebedev, F. Mathieu, F. de Montgolfier, J. Reynier, and L. Viennot. Acyclic preference systems in p2p networks. In *Euro-Par*, volume 4641 of *Lecture Notes in Computer Science*, pages 825–834. Springer, 2007.
- [7] A.-T. Gai, F. Mathieu, F. de Montgolfier, and J. Reynier. Stratification in p2p networks: Application to bittorrent. In *ICDCS*, page 30. IEEE Computer Society, 2007.
- [8] D. Gale and L. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69:9–15, 1962.
- [9] R. W. Irving, D. Manlove, and S. Scott. The hospitals/residents problem with ties. In *SWAT '00: Proceedings of the 7th Scandinavian Workshop on Algorithm Theory*, pages 259–271, London, UK, 2000. Springer-Verlag.
- [10] R. W. Irving and D. F. Manlove. The stable roommates problem with ties. *J. Algorithms*, 43(1):85–105, 2002.
- [11] J. Kleinberg. The small-world phenomenon: an algorithm perspective. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 163–170, New York, NY, USA, 2000. ACM.
- [12] D. Lebedev, F. Mathieu, L. Viennot, A.-T. Gai, J. Reynier, and F. de Montgolfier. On using matching theory to understand P2P network design. In *International Network Optimization Conference (INOC)*, 2007.
- [13] E. Lebhar, P. Fraigniaud, and L. Viennot. The inframetric model for the internet. In *Proceedings of the 27th IEEE International Conference on Computer Communications (INFOCOM)*, pages 1–9, 2008. To appear.
- [14] J. Marshall Ash, J. Cohen, C. Freiling, and D. Rinne. Generalizations of the wave equation. *Transactions of the American Mathematical Society*, 338(1):57–75, Jul 1993.
- [15] F. Mathieu. Upper bounds for stabilization in acyclic preference-based systems. In *SSS*, volume 4838 of *Lecture Notes in Computer Science*, pages 372–382. Springer, 2007.
- [16] F. Mathieu. Self-stabilization in preference-based systems. *Peer-to-Peer Networking and Applications*, 1(2):104–121, sept 2008.
- [17] S. Mertens. Random stable matchings. *J. Stat. Mech.: Theor. Exp.*, page P10008, 2005.
- [18] Meridian Project. <http://www.cs.cornell.edu/People/egs/meridian/>.
- [19] A. E. Roth. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, 92(6):991–1016, 1984.
- [20] A. E. Roth, T. Sonmez, and M. Utku Unver. Pairwise kidney exchange. *Journal of Economic Theory*, 125(2):151–188, December 2005. available at <http://ideas.repec.org/a/eee/jetheo/v125y2005i2p151-188.html>.
- [21] J. J. M. Tan. A necessary and sufficient condition for the existence of a complete stable matching. *J. Algorithms*, 12(1):154–178, 1991.

## A Proof of Theorem 1

The proof relies on the following steps:

- we prove that the  $\tilde{\mathcal{D}}_N$  functions are uniformly Cauchy on  $[0, 1]^2$ ;
- we use the Cauchy convergence to show that  $\mathcal{S}_N$  and  $\tilde{\mathcal{D}}_N$  have limits  $\mathcal{S}_\infty$  and  $\mathcal{D}_\infty$ , and we give a PDE verified by  $\mathcal{S}_\infty$ ;
- we solve the PDE, and use the solution to get  $\mathcal{D}_\infty$ .

### A.1 Uniform convergence

Let  $N$  be fixed, and  $N_1, N_2$  be two integers greater than  $N$ . The corresponding Erdős-Rényi probabilities are  $p_1 = \frac{d}{N_1}$  and  $p_2 = \frac{d}{N_2}$ . We consider the error function defined by

$$E(\alpha, \beta) = |\mathcal{D}_{N_1}(\alpha, \beta) - \mathcal{D}_{N_2}(\alpha, \beta)|. \quad (23)$$

For proving that  $\tilde{\mathcal{D}}_N$  is uniformly Cauchy, we need to find a bound for  $E$  that applies for any  $(\alpha, \beta) \in [0, 1]^2$ , and that tends towards 0 as  $N$  goes to infinity.

Let  $\alpha_1, \beta_1, \alpha_2, \beta_2$  be respectively  $\frac{\lfloor N_1 \alpha \rfloor}{N_1}, \frac{\lfloor N_1 \beta \rfloor}{N_1}, \frac{\lfloor N_2 \alpha \rfloor}{N_2}, \frac{\lfloor N_2 \beta \rfloor}{N_2}$ . Using (4) and (6), we have, for  $k \in \{1, 2\}$ ,

$$\begin{aligned} \tilde{\mathcal{D}}_{N_k}(\alpha, \beta) = & d(1 - \int_0^{\beta_k} \mathcal{D}_{N_k}(\alpha_k, x) dx) \times \\ & \times (1 - \int_0^{\alpha_k} \mathcal{D}_{N_k}(\beta_k, x) dx) \end{aligned}$$

It would be nice to have  $\alpha$  and  $\beta$  instead of  $\alpha_k$  and  $\beta_k$ , and  $\tilde{\mathcal{D}}$  instead of  $\mathcal{D}$ . In order to do that, we notice the following:

- $\mathcal{D}_{N_k}(\alpha_k, x) = \mathcal{D}_{N_k}(\alpha, x)$ ;
- same for  $\mathcal{D}_{N_k}(\beta_k, x)$ ;
- $S \leq 1$ , so we have  $\mathcal{D}_{N_k} \leq N_k p_k = d$ . As  $\alpha - \alpha_k < \frac{1}{N_k} \leq \frac{1}{N}$ , it follows that  $\int_{\alpha_k}^{\alpha} \mathcal{D}_{N_k}(\beta, x) dx \leq \frac{d}{N}$ ;
- the same with  $\alpha$  and  $\beta$  switched;
- $\tilde{\mathcal{D}}_{N_k}(\alpha, x)$  is bounded by  $d$  and only differs from  $\mathcal{D}_{N_k}(\alpha, x)$  for  $\lfloor N_k \alpha \rfloor = \lfloor N_k x \rfloor$ . It follows that  $|\int_0^{\beta} \tilde{\mathcal{D}}_{N_k}(\alpha, x) - \mathcal{D}_{N_k}(\alpha, x) dx| \leq \frac{d}{N}$ ;
- the same with  $\alpha$  and  $\beta$  switched.

We deduce that

$$\left| \int_0^{\beta_k} \mathcal{D}_{N_k}(\alpha_k, x) dx - \int_0^{\beta} \tilde{\mathcal{D}}_{N_k}(\alpha, x) dx \right| \leq \frac{2d}{N},$$

and the same with  $\alpha$  and  $\beta$  switched. Then, if we call

$$\tilde{\mathcal{S}}_{N_k}(\alpha, \beta) = 1 - \int_0^{\beta} \tilde{\mathcal{D}}_{N_k}(\alpha, x) dx, \quad (24)$$

we have

$$|\tilde{\mathcal{D}}_{N_k}(\alpha, \beta) - d\tilde{\mathcal{S}}_{N_k}(\alpha, \beta)\tilde{\mathcal{S}}_{N_k}(\beta, \alpha)| \leq \frac{8d^2}{N}. \quad (25)$$

This gives us

$$E(\alpha, \beta) \leq \frac{16d^2}{N} + d \left| \tilde{\mathcal{S}}_{N_1}(\alpha, \beta)\tilde{\mathcal{S}}_{N_1}(\beta, \alpha) - \tilde{\mathcal{S}}_{N_2}(\alpha, \beta)\tilde{\mathcal{S}}_{N_2}(\beta, \alpha) \right| \quad (26)$$

Using the definition of  $\tilde{\mathcal{S}}_{N_k}$ , we see that

$$\begin{aligned} & \left| \tilde{\mathcal{S}}_{N_1}(\alpha, \beta) \tilde{\mathcal{S}}_{N_1}(\beta, \alpha) - \tilde{\mathcal{S}}_{N_2}(\alpha, \beta) \tilde{\mathcal{S}}_{N_2}(\beta, \alpha) \right| \leq \\ & \int_0^\beta E(\alpha, x) dx + \int_0^\alpha E(\beta, x) dx + \\ & \left( \int_0^\beta E(\alpha, x) dx \right) \left( \int_0^\alpha E(\beta, x) dx \right) \end{aligned}$$

Note, that both  $\tilde{\mathcal{D}}_{N_1}(\alpha, \cdot)$  and  $\tilde{\mathcal{D}}_{N_2}(\alpha, \cdot)$  are probabilities, so we can bound  $\int_0^\beta E(\alpha, x) dx$  by 2 in the integral product. Then (26) becomes

$$E(\alpha, \beta) \leq \frac{16d^2}{N} + d \left( \int_0^\beta E(\alpha, x) dx + 3 \int_0^\alpha E(\beta, x) dx \right) \quad (27)$$

We now want to merge  $\alpha$  and  $\beta$  into a single variable. Therefore, we define  $F(\gamma) := \sup_{\alpha \leq 1, \beta \leq 1, \alpha + \beta \leq \gamma} E(\alpha, \beta)$ . For any  $\alpha \leq 1, \beta \leq 1, \gamma \geq \alpha + \beta$ , we have

$$\begin{aligned} \int_0^\beta E(\alpha, x) dx & \leq \int_0^\beta F(\alpha + x) dx \\ & \leq \int_0^{\alpha + \beta} F(x) dx \\ & \leq \int_0^\gamma F(x) dx, \end{aligned}$$

and the same for  $\int_0^\alpha E(\beta, x) dx$ . It follows that

$$F(\gamma) \leq \frac{16d^2}{N} + 4d \int_0^\gamma F(x) dx \quad (28)$$

It follows that  $F(\gamma) \leq \frac{16d^2}{N} e^{4d\gamma}$  by Grönwall's lemma [3]. As a special case, for all  $\alpha, \beta \leq 1$ , we have

$$E(\alpha, \beta) \leq F(2) \leq \frac{16d^2}{N} e^{8d}. \quad (29)$$

This concludes the proof that  $\tilde{\mathcal{D}}_N$  is uniformly Cauchy.

## A.2 PDE

As  $\tilde{\mathcal{D}}_N$  is uniformly Cauchy on  $[0, 1]^2$ , it converges towards a function  $\mathcal{D}_\infty$ . Using (24), we deduce that  $\tilde{\mathcal{S}}_N$  converges towards a continuous function  $\mathcal{S}_\infty$ , and that  $-\mathcal{D}_\infty$  is the partial derivative of  $\mathcal{S}_\infty$  with respect to its second variable.

Then, if we make  $N$  go to infinity in (25), we obtain the PDE verified by  $\mathcal{S}_\infty$ :

$$\partial_y \mathcal{S}_\infty(\alpha, \beta) = -d \mathcal{S}_\infty(\alpha, \beta) \mathcal{S}_\infty(\beta, \alpha), \quad (30)$$

with limit condition  $\mathcal{S}_\infty(\alpha, 0) = 1$ .

Notice that (30) proves that  $\mathcal{D}_\infty$  is continuous.

## A.3 Resolution

Note, that for  $\alpha = 0$ , (30) immediately gives  $\mathcal{S}_\infty(0, \beta) = e^{-d\beta}$ .

To go further, we introduce the auxiliary function  $f(\alpha, \beta) := \log\left(\frac{\mathcal{S}_\infty(\beta, \alpha)}{\mathcal{S}_\infty(\alpha, \beta)}\right)$ .

$f$  is skew-symmetric. Its first partial derivative is:

$$\begin{aligned} \partial_x f(\alpha, \beta) & = \frac{\partial_x \mathcal{S}_\infty(\alpha, \beta)}{\mathcal{S}_\infty(\alpha, \beta)} - \frac{\partial_y \mathcal{S}_\infty(\beta, \alpha)}{\mathcal{S}_\infty(\beta, \alpha)} \\ & = \frac{\partial_x \mathcal{S}_\infty(\alpha, \beta)}{\mathcal{S}_\infty(\alpha, \beta)} + d \mathcal{S}_\infty(\alpha, \beta) \end{aligned}$$

By differentiating again, we get the mixed derivative

$$\begin{aligned}
\partial_{xy}f(\alpha, \beta) &= \frac{\partial_{xy}\mathcal{S}_\infty(\alpha, \beta)}{\mathcal{S}_\infty(\alpha, \beta)} - \frac{\partial_x\mathcal{S}_\infty(\alpha, \beta)\partial_y\mathcal{S}_\infty(\alpha, \beta)}{(\mathcal{S}_\infty(\alpha, \beta))^2} \\
&\quad + d\partial_y\mathcal{S}_\infty(\alpha, \beta) \\
&= \frac{\partial_{yx}\mathcal{S}_\infty(\alpha, \beta)}{\mathcal{S}_\infty(\alpha, \beta)} + \frac{\partial_x\mathcal{S}_\infty(\alpha, \beta)d\mathcal{S}_\infty(\beta, \alpha)}{\mathcal{S}_\infty(\alpha, \beta)} \\
&\quad + d\partial_y\mathcal{S}_\infty(\alpha, \beta) \\
&= -\frac{d\partial_x\mathcal{S}_\infty(\alpha, \beta)\mathcal{S}_\infty(\beta, \alpha)}{\mathcal{S}_\infty(\alpha, \beta)} \\
&\quad - \frac{d\mathcal{S}_\infty(\alpha, \beta)\partial_y\mathcal{S}_\infty(\beta, \alpha)}{\mathcal{S}_\infty(\alpha, \beta)} \\
&\quad + d\frac{\partial_x\mathcal{S}_\infty(\alpha, \beta)\mathcal{S}_\infty(\beta, \alpha)}{\mathcal{S}_\infty(\alpha, \beta)} + d\partial_y\mathcal{S}_\infty(\alpha, \beta) \\
&= 0 \quad (= \partial_{yx}f(\alpha, \beta))
\end{aligned}$$

The only global solutions to the wave equation  $f_{xy} = 0$  are those of the form  $f(\alpha, \beta) = a(\alpha) + b(\beta)$  (see [14], for instance). Given that  $f$  is skew-symmetric, the solution is indeed of the form  $f(\alpha, \beta) = a(\alpha) - a(\beta)$ . The border conditions immediately give  $f(\alpha, \beta) = d(\beta - \alpha)$ .

We deduce  $\mathcal{S}(\beta, \alpha)$ :  $\mathcal{S}(\beta, \alpha) = \mathcal{S}(\alpha, \beta)e^{d(\beta-\alpha)}$ .

If we treat  $\mathcal{S}_\infty$  as a function of  $\beta$  with  $\alpha$  as parameter, equation (30) becomes

$$\dot{\mathcal{S}}_\alpha(\beta) = -d\mathcal{S}_\alpha^2(\beta)e^{d(\beta-\alpha)} \quad (31)$$

From there, one get  $\mathcal{S}_\infty(\alpha, \beta) = \frac{1}{K+e^{d(\beta-\alpha)}}$ . Given that  $\mathcal{S}_\infty(\alpha, 0) = 1$ , the solution is:

$$\mathcal{S}_\infty(\alpha, \beta) = \frac{1}{1 - e^{-d\alpha} + e^{d(\beta-\alpha)}}.$$

Using  $\mathcal{D}_\infty = -\partial_y\mathcal{S}_\infty$ , one get (8). This concludes the proof of Theorem 1.

## B Exact resolution of the node-based stable configuration

### B.1 Recursive formula

For  $b = 1$ , we can give an explicit recursive formula for  $D(i, j)$ . The first step is to compute  $D(1, k)$ , for  $2 \leq k \leq n$ . As 1 is the best node, it can choose the best of its neighbors, so  $D(1, k)$  is the probability that  $k$  is the best of 1's neighbors. In other words, this is the probability that  $k$  is acceptable for 1, while all nodes  $l$  with  $1 < l < k$  are not. This gives us

$$D(1, k) = p(1 - p)^{k-2}. \quad (32)$$

Now, we consider two nodes  $i$  and  $j$  such that  $1 < i < j \leq n$ .  $D(i, j) = P(i \leftrightarrow j)$  can be calculated with a proper conditioning on the mate  $k$  (if any) of 1. The key is to notice that if 1 is mated with  $k$ , the both of them can be virtually removed from the graph. The remaining graph is still Erdős-Rényi and the probabilities are the same up to a slight relabeling:

- if  $k = i$  or  $k = j$ , then  $i$  cannot be mated with  $j$ ;
- if  $1 < k < i$ ,  $i$  and  $j$  can be virtually relabeled  $i - 2$  and  $j - 2$  (cf Figure 7a), so we have  $P(i \leftrightarrow j | 1 < k < i) = P((i - 2) \leftrightarrow (j - 2))$ ;
- if  $i < k < j$ ,  $i$  and  $j$  can be virtually relabeled  $i - 1$  and  $j - 2$  (cf Figure 7b), so we have  $P(i \leftrightarrow j | i < k < j) = P((i - 1) \leftrightarrow (j - 2))$ ;
- if 1 is not mated or  $k > j$  (notation:  $k \not\leq j$ ),  $i$  and  $j$  can be virtually relabeled  $i - 1$  and  $j - 1$  (cf Figure 7c), so we have  $P(i \leftrightarrow j | k \not\leq j) = P((i - 1) \leftrightarrow (j - 1))$ .

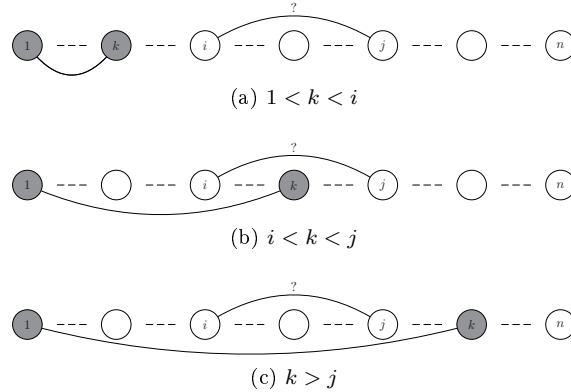


Figure 7: Using the mate of 1 to deduce  $D(i, j)$ .

Under this conditioning, we get

$$\begin{aligned} D(i, j) = & P(i \leftrightarrow j | 1 < k < i)P(1 < k < i) \\ & + P(i \leftrightarrow j | i < k < j)P(i < k < j) \\ & + P(i \leftrightarrow j | k \not\leq j)P(k \not\leq j). \end{aligned} \quad (33)$$

This leads to the following formula for  $D$ :

$$\begin{aligned} D(i, j) = & (1 - S(1, i))D(i - 2, j - 2) \\ & + (S(1, i + 1) - S(1, j))D(i - 1, j - 2) \\ & + S(1, j + 1)D(i - 1, j - 1). \end{aligned} \quad (34)$$

From (32), we have  $S(1, k) = (1 - p)^{k-2}$ . This gives

$$\begin{aligned} D(i, j) = & A(i)D(i - 2, j - 2) + B(i, j)D(i - 1, j - 2) + C(j)D(i - 1, j - 1), \text{ with} \\ A(i) = & 1 - (1 - p)^{i-2} \\ B(i, j) = & (1 - p)^{i-1} - (1 - p)^{j-2} \\ C(j) = & (1 - p)^{j-1} \end{aligned} \quad (35)$$



Now, in order to give a fluid limit, it can be convenient to reduce 34 to an expression of the complementary cumulative distribution  $S$ . Using the definition  $S(i, j) = \sum_{l \leq j} D(i, l)$ , Equation 35 becomes, after simplification,

$$\begin{aligned} S(i, j) &= A(i)S(i-2, j-2) + B(i, j)S(i-1, j-2) + C'(j)S(i-1, j-1), \text{ with} \\ C_2(j) &= (1-p)^{j-2}. \end{aligned} \quad (36)$$

## B.2 Uniform convergence

Like for the mean formula, we can prove that the scaling  $\mathcal{D}_N(\alpha, \beta) = ND(\lfloor N\alpha \rfloor + 1 + 1, \lfloor N\beta \rfloor + 1 + 1)$  is uniformly Cauchy. The sketch of proof is the same: clean the boundary of the integrals and the other  $O(\frac{1}{N})$  offsets, then use an auxiliary error variable  $\gamma$  and use Grönwall's lemma to conclude. This guarantees the convergence of  $\mathcal{D}_N$  and  $\mathcal{S}_N$ .

## B.3 PDE

We will use the fact that if we use the scaling  $i \rightarrow \lfloor N\alpha \rfloor + 1, j \rightarrow \lfloor N\beta \rfloor + 1$ , then

- $A(i)$  converges towards  $1 - e^{-d\alpha}$ ,
- $B(i, j)$  converges towards  $e^{-d\alpha} - e^{-d\beta}$ ,
- $C(j)$  and  $C_2(j)$  both converge towards  $e^{-d\beta}$ .

The first step is to *translate* (36) into an expression of  $\mathcal{S}_N$ : with  $\alpha = \frac{i-1}{N}$  and  $\beta = \frac{j-1}{N}$ , we obtain

$$\mathcal{S}_N(\alpha, \beta) = A(i)\mathcal{S}_N(\alpha - \frac{2}{N}, \beta - \frac{2}{N}) + B(i, j)\mathcal{S}_N(\alpha - \frac{1}{N}, \beta - \frac{2}{N}) + C_2(j)\mathcal{S}_N(\alpha - \frac{1}{N}, \beta - \frac{1}{N}). \quad (37)$$

We notice that  $(A + B + C)(i, j) = 1 - p(1-p)^{i-2}$ . If we remove  $(A + B + C)(i, j)\mathcal{S}_N(\alpha, \beta - \frac{2}{N})$  from each side of (36), and multiply the result by  $N$ :

- the left part becomes

$$N(\mathcal{S}_N(\alpha, \beta) - \mathcal{S}_N(\alpha, \beta - \frac{2}{N})) + d(1-p)^{i-2}\mathcal{S}_N(\alpha, \beta - \frac{2}{N}),$$

which converges as  $N \rightarrow \infty$  towards

$$2\frac{\partial \mathcal{S}_\infty}{\partial \beta} + de^{-d\alpha}\mathcal{S}_\infty,$$

- the right part becomes

$$A(i)N(\mathcal{S}_N(\alpha - \frac{2}{N}, \beta - \frac{2}{N}) - \mathcal{S}_N(\alpha, \beta - \frac{2}{N})) + (B(i, j) + C(j))N(\mathcal{S}_N(\alpha - \frac{1}{N}, \beta - \frac{2}{N}) - \mathcal{S}_N(\alpha, \beta - \frac{2}{N})) + C(j)N(\mathcal{S}_N(\alpha - \frac{1}{N}, \beta - \frac{1}{N}) - \mathcal{S}_N(\alpha - \frac{1}{N}, \beta - \frac{2}{N})),$$

which converges as  $N \rightarrow \infty$  towards

$$-2(1 - e^{-d\alpha})\frac{\partial \mathcal{S}_\infty}{\partial \alpha} - e^{-d\alpha}\frac{\partial \mathcal{S}_\infty}{\partial \alpha} + e^{-d\beta}\frac{\partial \mathcal{S}_\infty}{\partial \beta}.$$

So after scaling, the recursive equation is now:

$$2\frac{\partial \mathcal{S}_\infty}{\partial \beta} + de^{-d\alpha}\mathcal{S}_\infty = -2(1 - e^{-d\alpha})\frac{\partial \mathcal{S}_\infty}{\partial \alpha} - e^{-d\alpha}\frac{\partial \mathcal{S}_\infty}{\partial \alpha} + e^{-d\beta}\frac{\partial \mathcal{S}_\infty}{\partial \beta}.$$

In other words,  $\mathcal{S}$  verifies the PDE:

$$(2 - e^{-d\alpha})\frac{\partial \mathcal{S}_\infty}{\partial \alpha} + (2 - e^{-d\beta})\frac{\partial \mathcal{S}_\infty}{\partial \beta} + de^{-d\alpha}\mathcal{S}_\infty = 0.$$

**Theorem 3** With the border condition  $\mathcal{S}_\infty(0, \beta) = e^{-d\beta}$ , the unique solution of this PDE is  $\mathcal{S}_\infty = \frac{1}{1 - e^{-d\alpha} + e^{d(\beta-\alpha)}}$ . The scaled version of  $D$ , denoted  $\mathcal{D}_\infty$ , thus verifies:

$$\mathcal{D}_\infty(\alpha, \beta) = -\frac{\partial \mathcal{S}_\infty}{\partial \beta}(\alpha, \beta) = \frac{de^{d(\beta-\alpha)}}{(1 - e^{-d\alpha} + e^{d(\beta-\alpha)})^2}.$$

**Proof:** Let us change the variables: put  $x = e^{d\alpha}$  and  $y = e^{d\beta}$ . We also make the PDE more symmetric by multiplying by  $x$ . Define  $u$  by putting  $\mathcal{S}_\infty(\alpha, \beta) = xu(x, y)$ . The PDE then becomes:

$$(2 - \frac{1}{x})(dx^2 \frac{\partial u}{\partial x} + dxu) + (2 - \frac{1}{y})xdy \frac{\partial u}{\partial y} + d\frac{1}{x}xu = 0.$$

ie:

$$\begin{cases} (2x - 1)\frac{\partial u}{\partial x} + (2y - 1)\frac{\partial u}{\partial y} + 2u = 0 \\ u(1, y) = \frac{1}{y} \end{cases}$$

This equation is a non-linear first order PDE:  $F(Du, u, x) = 0$ , where  $F$  is linear. To solve this PDE, we use the classical method of *characteristics* described in [5], chapter 3. Let  $X(s) = (x(s), y(s))$  ( $s$  in an interval of  $\mathbb{R}$ ), be a trajectory in the base space; define  $p(s) = Du(X(s))$  and  $z(s) = u(X(s))$ . Then, solving the equation  $F(p(s), z(s), S(s)) = 0$  leads to the equivalent system of ODE (we forget about  $p(s)$ , which is not required to solve the PDE with boundary condition, see [5] p 100 for further precisions):

$$\begin{cases} \dot{x}(s) = 2x(s) - 1 \\ \dot{y}(s) = 2y(s) - 1 \\ \dot{z}(s) = -2z(s) \\ z_0 = z(x_0 := 1, y_0) = \frac{1}{y_0} \end{cases}$$

where  $\dot{\cdot}$  stands for  $\frac{d}{ds}$ .

These 3 ODEs are with separable variables (Cauchy-Lipschitz theorem applies for existence and unicity). The solution with the boundary condition at  $s = 0$ ,  $x_0 = 1$ ,  $y_0 \in \mathbb{R}$ ,  $z_0 = \frac{1}{y_0}$  is:

$$\begin{cases} 2x(s) - 1 = e^{2s} \\ 2y(s) - 1 = (2y_0 - 1)e^{2s} \\ z(s) = \frac{1}{y_0}e^{-2s} \end{cases}$$

Now given  $(x, y)$ , we deduce  $s$  such that  $x(s) = x$  and  $y(s) = y$  then  $y_0$  and  $z(s) = u(x, y)$ :  $2y_0 - 1 = \frac{2y-1}{2x-1}$  then

$$u(x, y) = \frac{1}{y_0} \frac{1}{2x-1} = \frac{1}{x+y-1}.$$

Replacing  $x$  and  $y$  by  $e^{d\alpha}$  and  $e^{d\beta}$  concludes the proof.  $\square$



---

Unité de recherche INRIA Rocquencourt  
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes  
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique  
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

---

Éditeur  
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399