Selecting an optimal number of particles to fit large network computational models with random PSO

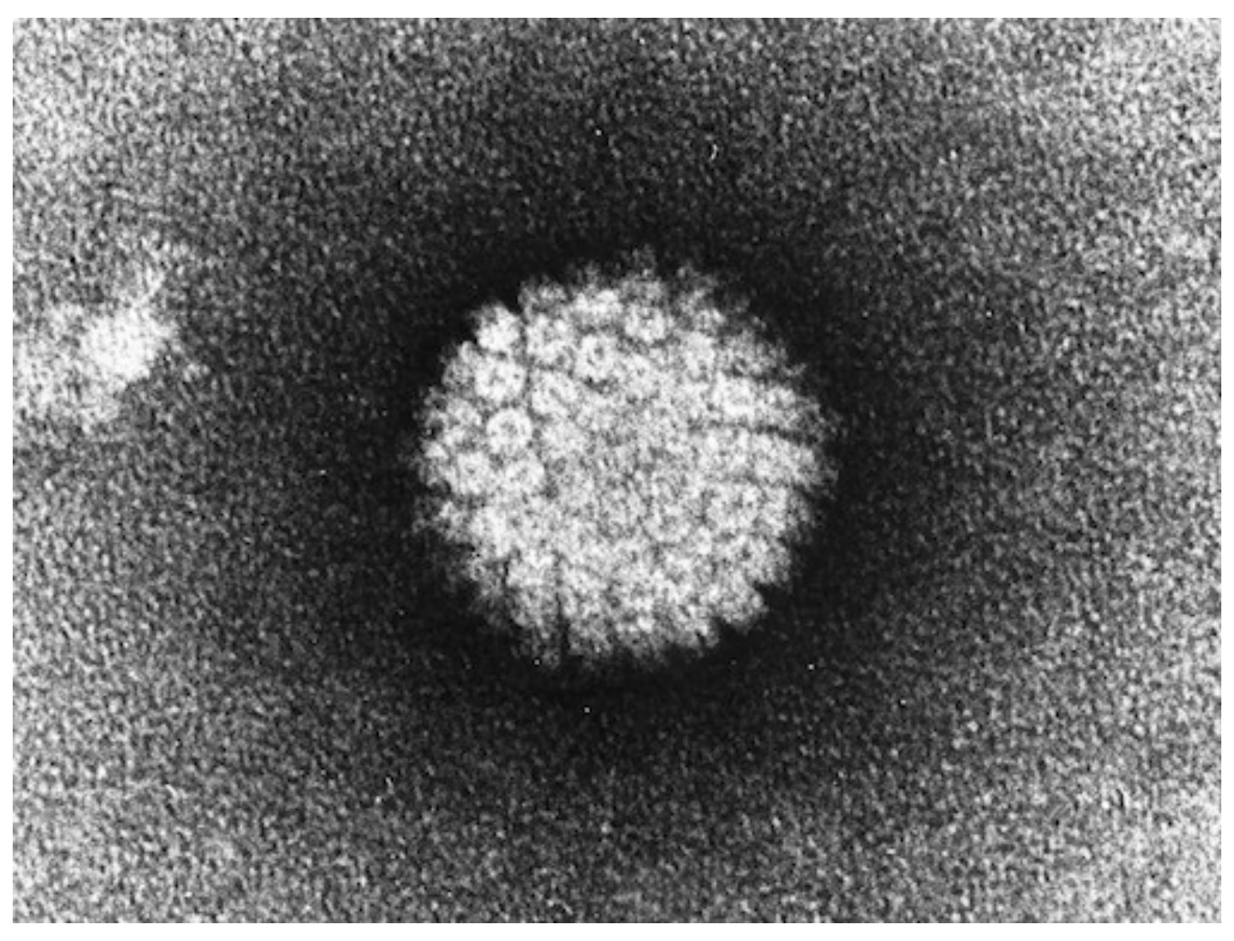
José-Ignacio Hidalgo[1], Clara Burgos[2], David Martínez-Rodríguez[2], Rafael J. Villanueva[2]

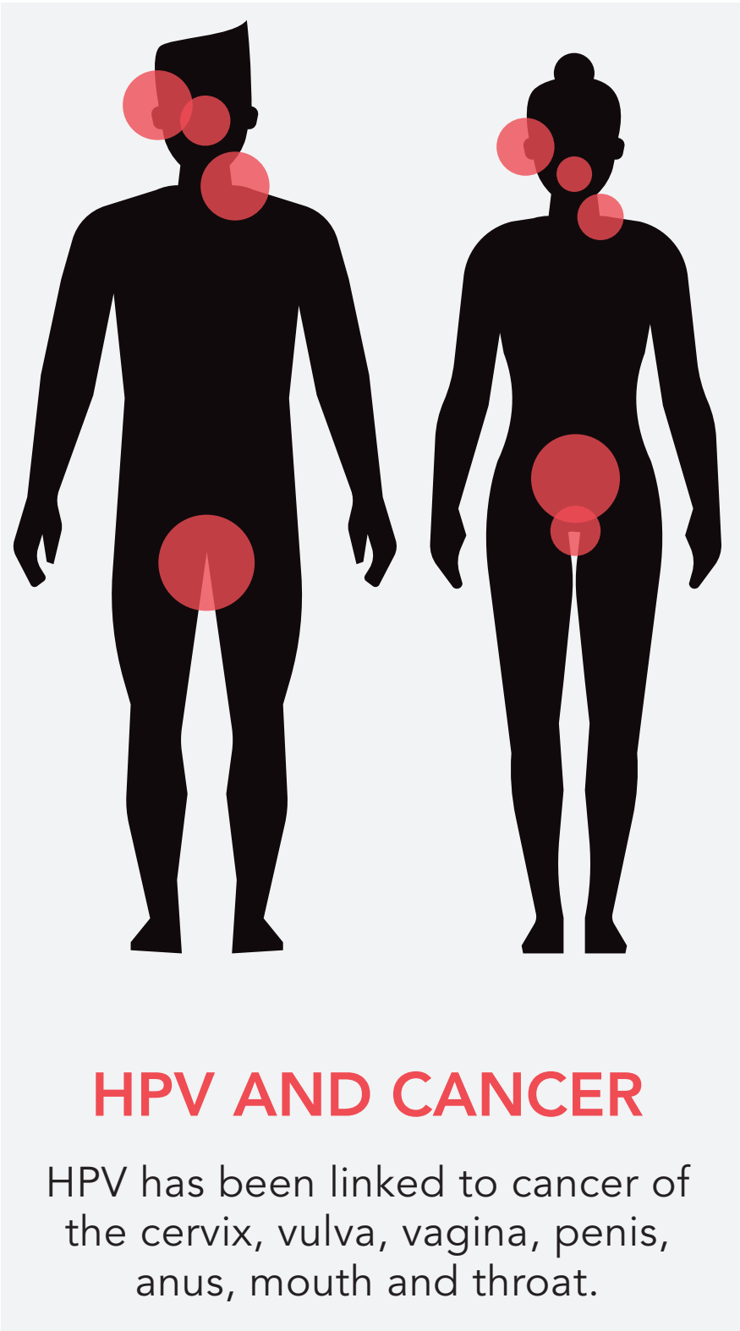
hidalgo@dacya.ucm.es, clabursi@posgrado.upv.es, damarro3@etsii.upv.es, rjvillan@imm.upv.es

1. Departamento de Arquitectura de Computadores y Automática, Universidad Complutense de Madrid, Madrid, Spain.
2. Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de Valéncia, Valencia, Spain.

Introduction

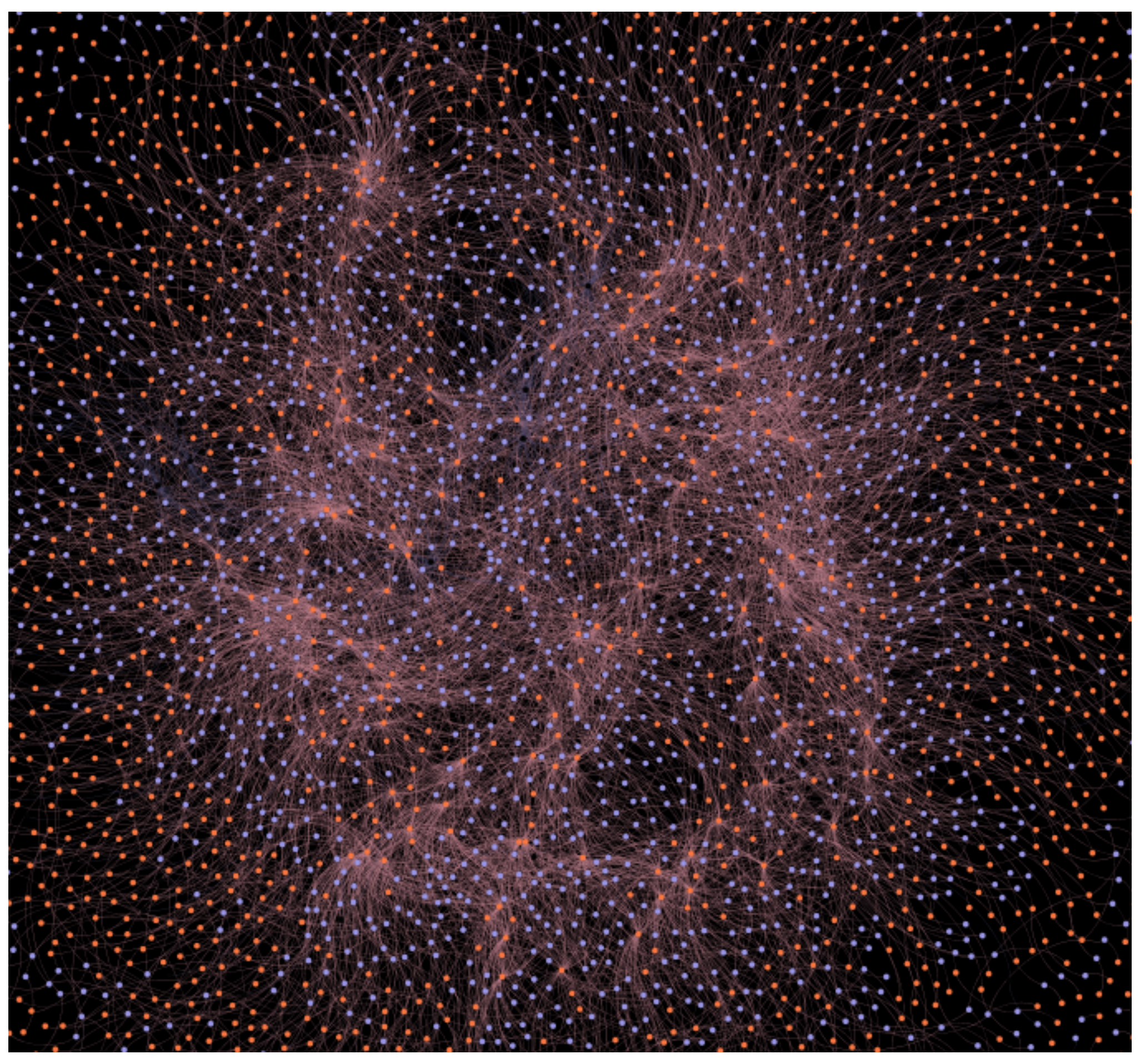
Human papillomaviruses (HPV) include more than 100 genotypes of viruses that infect cutaneous, genital and respiratory epithelia of humans. It is transmitted via vaginal, anal or oral sex with someone who has the virus. In general, it is asymptomatic, inoffensive and disappears spontaneously, however, if it persists, may develop anogenital warts, secondarily juvenile onset of recurrent respiratory papillomatosis, cervical cancer, other anogenital cancers and head and neck cancers.





To describe accurately the HPV transmission dynamics, we proposed the building of networks of lifetime sexual partners (LSP).

These networks are built to reproduce the demography and the sexual habits in Spain, for men, women and men who have sex with men (MSM).



Then, we describe the HPV transmission dynamics over the LSP network, it is necessary to include 11 model parameters that will govern the HPV contagion evolution.

These model parameters are mainly related to the duration of the infection and the probability of contagion, for men and women and depending on the HPV type.

Therefore, we want to find the model parameters in such a way the computational model output is close to the data of prevalence of HPV in Spain.

Goal

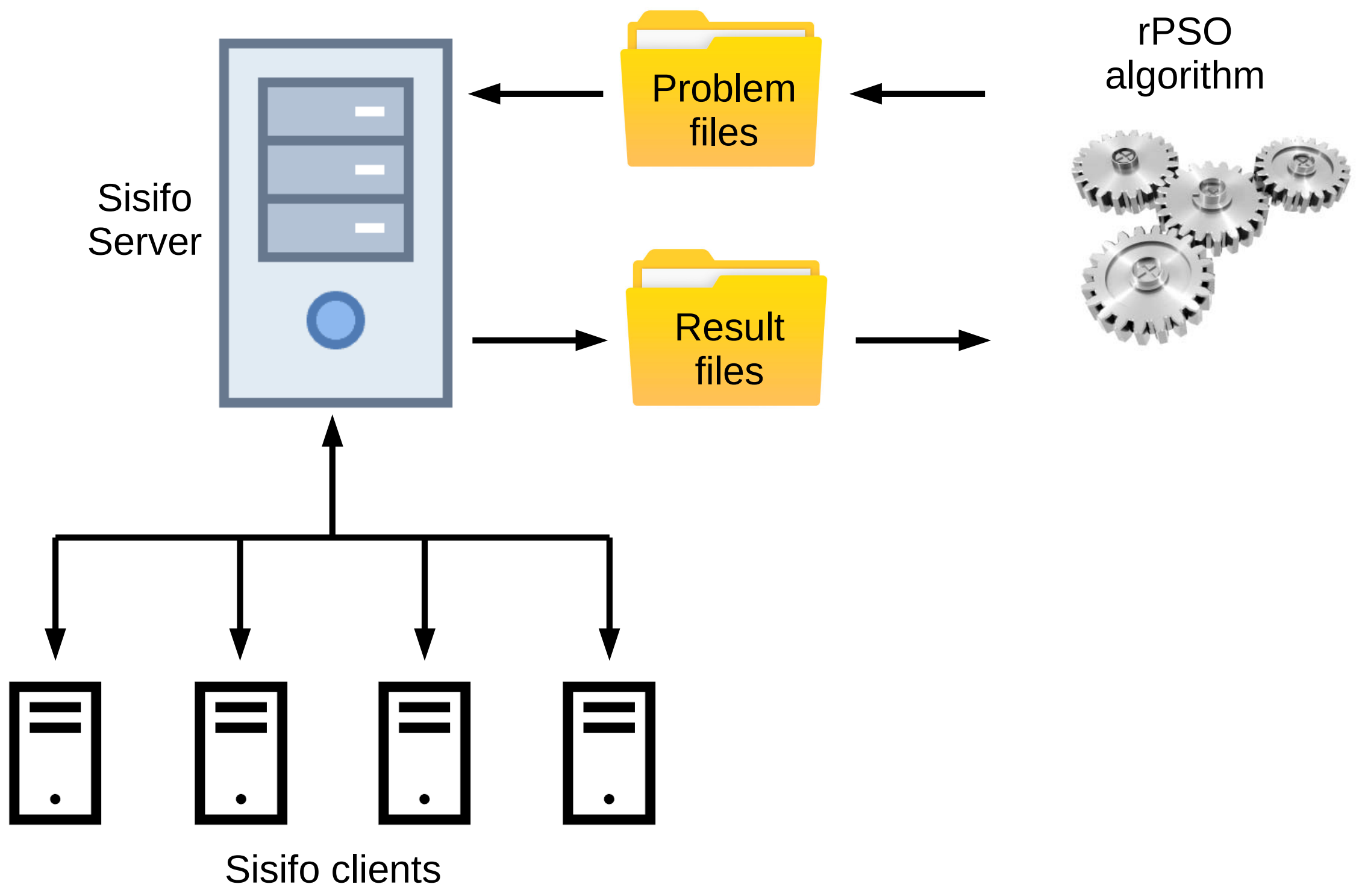
In order to the model calibration with HPV prevalence data, we are going to use the Particle Swarm Optimization (PSO) algorithm in a distributed computing environment.

Thus, the objective of the present study is to determine the best implementation of the PSO, in terms of quality of the solutions and computational efficiency.

Experiment preparation

To perform the model calibration, we propose the inclusion of an asynchronous version of the Random Particle Swarm Optimization (rPSO) in a distributed computing paradigm named Sisifo, a client-server based system designed to solve problems using distributed computers. Sisifo is able to assign tasks to a set of processors, wait for the tasks to be completed and collect the results for further analysis.

The asynchronous version of rPSO works as follows: when the fitness of a particle has been evaluated, this particle is immediately updated without waiting for the evaluation of the remainder particles, considering the current existing global best and its individual best particles.



rPSO manages the generation of new problems from the new particles and put them in the *Problem files* folder and the reading and processing of the solution files located in the *Result files* folder.

Design and results of the experiment

We can not affirm that the higher the number of particles, the better the quality of the solution. Also, the use of more processors does not means lower execution times.

Then, the question is, what is the quality of the solutions for different number of particles?

Table 1 shows important information regarding the quality of the solutions for different number of particles and five runs of each configuration. Each run of rPSO carried out 1200 particles evaluations.

# Part.	Total #0s	Avg #0s	Avg First	Best at end	Worst at end	Avg at end	Total Time	Avg Worst t	Avg Best t
25	18	3.60	**531.00**	**0.015486**	**0.429090**	**0.122573**	8422.00	**148.60**	**107.00**
30	**28**	**5.60**	594.00	0.003785	0.445587	0.131193	**6908.00**	186.60	112.00
35	8	1.60	834.80	0.010211	0.576688	0.149215	7105.00	250.60	119.60
40	16	3.20	692.80	0.003332	0.524417	0.132628	8808.60	389.40	144.60
45	14	2.80	790.40	0.006458	0.639775	0.149670	10004.00	502.00	177.20
50	11	2.20	796.00	0.005168	0.658020	0.139089	11124.40	648.40	220.00
55	2	0.40	1171.20	0.004330	0.518894	0.137088	11905.80	775.00	207.40
60	12	2.40	765.00	0.002153	0.546259	0.132464	13787.80	857.00	245.40
64	5	1.00	985.40	0.002763	0.551228	0.140977	18770.00	1041.20	379.20

Table 1: Analysis of the quality and execution times of different rPSO configurations. Results of 1200 evaluations for different number of particles on the rPSO process (# Part.).

Red color indicates the worst configuration, **bold** letters indicate the best and blue color the second best configuration.

Avg First: the iteration at which the first zero appears on average.

Best at end, Worst at end and Avg at end: from the last population of particles, Fitness of the best, worst and average, averaged over five runs.

Simulations performed on a computer with 64 cores on 8 Xeon Sandy Bridge E5-4620 running at 2,2 Ghz, with 16 MB of cache memory and 512 GB RAM memory.

Conclusion

25 and 30 particles are the preferred configurations, since we obtained the higher number of zeros in total and on average and the lower executions times.

Despite the results of 25 particles show a very good run with 7 zeros and also a very good execution time, we can conclude that the configuration with 30 particles should be selected bearing in mind both quality and execution time.

Results on total number of 0s and total time were statistically significant with p value of 0.1 after an ANOVA analysis.