SUMMER UNIVERSITY COORDINATION TEAM

Preselection Algorithm (non-IT explanation)

Author: Raúl Avilés Poblador (SUCT, AEGEE-Castelló)

May 6th, 2011

Introduction: a good algorithm for applicants and SU organisers

This preselection algorithm was explained during the IT Progress Meeting at Spring Agora Leiden 2010 and during a workshop at NWM Valencia 2010. This document will try to explain the preselection algorithm that has been used in the SU Project 2010 and 2011 as simply as possible (non-IT explanation). The source code is public and has been shared among the ITC.

The application of each member consists of: each applicant can fill a unique application, where can select up to three SU in order of preference.

The aim is that people who have selected more than one SU (i. e. two or three) get preselected to the SU where they have more possibilities to be chosen.

Therefore, the aim of the preselection algorithm is to weight the "worth" of each application and preselect each member, when possible, to the choice where he/she has more possibilities to be chosen, i. e., to the SU where the most balanced group of applications is. The position (1, 2 or 3) of the SU choices in the application is also taken into account.

We assume that locals will always look for a balanced group of participants, therefore our criteria are: gender, age, nationality and antenna.

This way we give the applicant what he/she wants (more possibilities to be chosen and consideration for the choices of preference) and we also give the locals what they want (a balanced group of applications).

Everything, of course, is subordinated to the nature of the applications. This means that if there are 80 applications of women and 40 applications of men to one SU, after preselection the percentages will still be kind of similar (and the applications would be for example around 50 women - 25 men). The preselection algorithm doesn't make miracles.

Structure of the preselection algorithm

The preselection algorithm consists of many technical steps, but I will try to explain them all only in few steps, taking them as conceptual steps.

Step 1

Applicants who only applied to one SU are automatically preselected to this SU.

Step 2

Only applications with more than one SU choice proceed to the second step. At this step each application is split in subapplications depending on how many choices this applicant has made. Each subapplication is weighted according to these criteria:

Criterion	Points
option1	4
option2	3.25
option3	2.5
gender	up to 7
age	~ can vary
nationality	up to 5
antenna	up to 0.5

Each application has, among others, these fields:

- name
- option1: SU choice 1
- option2: SU choice 2
- option3: SU choice 3
- gender
- age
- nationality
- antenna
- sc1, sc2 and sc3: score 1, 2 and 3.

How is the punctuation for each subapplication calculated?

The application will be processed as 3 “subapplications” (3 or 2, depending on how many options did he/she choose.)

The system has to calculate the scores sc1, sc2 and sc3. Each of this scores means the “quality” of this subapplication depending on the variables gender, age, nationality, antenna, and which option it is (1, 2 or 3). To see how it works, let's make an example:

weight	-	4	3.25	2.5	7	~	5	0.5			
	name	option1	option2	option3	gender	age	nationality	antenna	sc1	sc2	sc3
	Raul	Peiraiás	Ljubljana	Passau	male	22	Spain	Castelló			

Let's imagine I'm applying for 3 SUs in this order of preference, and my application data is the one in the table above.

For calculating “sc1” (score1, that corresponds to the “quality of my application to the SU of Peiraiás):

First of all, the system makes a general count (imagine that the numbers in parenthesis are the imaginary counts):

- How many people applied to Peiraiás SU? (in any of the preferences order) (107)
- How many women applied? (72)
- How many men applied? (35)
- Which is the average age of the applicants? (23)
- How many people of my own country (Spain) also applied to this SU? (18)
- How many people of my own antenna (AEGEE-Castelló) also applied to this SU? (2)

Therefore, the calculation of the sc1 is based on the "weights" of the above variables.

Choice points

Peiraiás SU is my first option, so: + 4

Gender points

Calculation is: $\left(1 - \frac{number\ of\ men\ that\ applied}{total\ number\ of\ people\ that\ applied}\right) * 7$

I'm a man, so: $\left(1 - \frac{35}{107}\right) * 7 = +4.71$

Age points

I calculate how far is my age from the average. The further, the more punctuation (this means more balanced ages in the groups, however it is possible that people in the middle gets less punctuation and people in extremes gets more). It could have been made in other way, but to simplify we do it like this.

I subtract my age from the average (in absolute value), and I divide it between 5 to have a nice number.

Calculation is: $\frac{|average\ age\ in\ the\ SU - my\ age|}{5}$

I'm 22 and the average in Peiraiás SU is 22, so: $\frac{|23-22|}{5} = +0.2$

Nationality points

Calculation is: $\left(1 - \frac{number\ of\ applicants\ from\ my\ country}{total\ number\ of\ people\ that\ applied}\right) * 5$

22 people applied from Spain, so: $\left(1 - \frac{18}{107}\right) * 5 = +4.16$

Antenna points

I weight it relative to the people that applied from my own country.

Calculation is: $\left(1 - \frac{number\ of\ applicants\ from\ my\ antenna}{number\ of\ applicants\ from\ my\ country}\right) * 0.5$

2 people applied from AEGEE-Castelló, so: $\left(1 - \frac{2}{18}\right) * 0.5 = +0.44$

Total points

If put everything together:

$$\boldsymbol{sc1} = 4 + 4.71 + 0.2 + 4.16 + 0.44 = \mathbf{13.51}$$

Therefore, 13.51 is the score of my subapplication to Peiraiás SU.

The algorithm does the same for the other two "subapplications" (Ljubljana SU and Passau SU).

Step 3

Once all the scores of all the subapplications of all the applicants have been calculated, comes the third part of the algorithm.

Now all the SUs are filled with applications until have reached the total number of places offered.

The criteria for filling the SUs is: the SU with less applications starts filling, and takes all the “subapplications” (obviously, subapplications to this SU) with higher scores, until reaching the number of places offered. Then the following SU with more subapplications, etc.

I start from the SUs with fewer applicants than places because this way they get the maximum applicants they can take.

Now, a balanced group with the exact number of places offered has been assured for each SU (the balancing has been made through the scores.)

In this step, each SU has the same number of applicants assigned than the number of places offered, except for the SUs where there are less applicants than places.

Step 4

Managing overbooking. Here, starting from SUs with more applicants (and more applicants than places), I ensure that the overbooking is at least of 100%. That means if a SU has 25 places, it should get at least 50 assigned applicants (if there are enough applicants to do that). Once the 100% overbooking has been reached, the overbooking is increased until all applicants are assigned to a SU. Again, the assignation of applicants is from subapplications with higher scores to subapplications with lower scores.

Miracles doesn't exist

The algorithm has tried to provide you with a number of applications that is at least double of the places you offer. At the same time, it has tried to give you a balanced group of participants in terms of gender, nationality, antennae and ages.

But, please, take into consideration that although the algorithm has tried to do its best, the result it is always up to the applicants’ characteristics and to the nature of our network.

Examples: usually there are more women' applications, therefore the preselection result will be always like that, because it is not due to the algorithm. The same happens with nationalities, etc.