methods\_desc

### Methods description

As shown by …, the yule distribution is not a good fit to CD sales data.

We propose a different approach.

1. For each agent randomly select a vertice of the graph (artist). Increment the playcount for each artist by number of agents who have chosen it.
2. In next steps make each agent do one of the following actions:
   * With probability 1-p, go to random vertice, with probabilities proportional to previous playcounts for given vertice. This is the same as the “snowball” part in yule process.
   * With probability 1-p, go to neighboring vertice (similar artist) to previously chosen one. If the artist from previous step does not have any neighboring vertice, go to random one (without weighting).
3. For each vertice selected, increment the playcount for each artist by number of agents who have chosen it.

This simulation is designed to test if the recommendation system provided by last.fm site is contribution to some artists being exceptionally popular.

Results of the simulations were then compared to empirical distribution of playcounts. Distributions were compared using quantile-quantile plots, empirical distribution being the base one. As a reference, two theoretical distributions were fitted using maximum-likelihood estimation. These were Yule distribution and Pareto distribution. Both of them are referenced by … for obtaining reasonable estimation of stardom distribution.

More formal tests of goodnness of fit to the dataset were also provided. A widely popular choice for comparing two arbitrary distributions is a Kolmorogov-Smirnof test. However, for this particular dataset important information is contained in the upper tail. For such distributions, Anderson-Darling test is more sufficient, and thus was also used.

The yule distribution arises from an underlying snowball process, which is defined as follows:

Fitting the yule distribution using MLE does not give any information about underlying p in this case. To obtain the estimate, we have simulated the process using … agents and 9998 artists (same number as in empirical data).

Sprawdzić generalized Yule?