pairing algorithm

We have looked at a lot of things. We've looked at the cosine simularity, the adjusted cosine simularity, the dot-product, the euclidiean distance and the pearsons correlation. Personally, I think that the pearsons correlation function produces the best results becuase it only takes into account the movies which both user A and B have watched. There is a problem with this approach: when a user hasn't watched a lot of movies, the movies which the user has watched make a bigger impact on the matching function. This means that the random factor involved in the rating process has a higher impact. For our system, we ignore this random factor since this is something which hopefully will be solved when the user rates more movies.

There is also the part where there is an odd number of users. When this happens, we add the last user to the group with which the user has the most correlation. This won't be a problem because when user A has a high correlation with user B en user C, user C en B probably also have a high correlation since they're both matched with user A