a. I would choose to use cosine distance for this data because the cosine is calculated by taking a dot product. If our test person hasn't seen a movie and someone we are trying to compare to has, then that doesn't have any effect on the dot product computation (and by extension the cosine distance). This would not be true for L1 or L2 distance: seeing a movie (and giving it any rating) would increase the distance from someone who has not seen it. Since this is an effect I would prefer to avoid, I would focus on the cosine distance as my metric of choice.
b. The cosine distance should handle this fine – this will result in a dot product adding up only terms associated with movies that have been seen by both parties we are trying to compute the distance between. This makes sense as most people have not rated most movies: we would only want to take these overlaps into account.