Making Matches - Recommending the right personality

Paul Schweiger

2103468

Multimedia Seminar II

Technische Universitt Darmstadt

Abstract—Recommender Systems can be used to guide our limited time and attention towards meaningful actions, items, content or people. To tailor results to a user, Recommender Systems need to account for a user's preferences, skills and personality and incorporate these into a unified user model. These models needs to be descriptive without overcomplicating the system. In social situations such as learning, recommendations become more complicated, as several users and their respective models need to be combined. Reciprocal recommendations can improve the quality of matchups by ensuring a good fit and benefit for both parties.

A recent approach to implement a reciprocal peer recommendation platform for learning purposes aims to show how these theories can be used in praxis. The system enables students to find study-partners of matching skill for different learning topics by reciprocally recommending them to each other according due to their respective preferences.

Based on this study, the intricacies of proper user modeling, reciprocal recommendations and researching new algorithms are highlighted. Common issues include unreliable self-reported user data, unevaluated algorithms and unfitting user models. Proper theoretical founding and preliminary considerations can solidify and improve reciprocal Recommender Systems.

I. Introduction

The modern world provides a plethora of opportunities, topics, people or products to consume, engage with or discuss, while people have a limited amount of attention and time to spend. As the offline world embraces online opportunities, this surplus of possible interactions is multiplied even further. In order to help us guide our attention and resources, basically every digital product tries to recommend meaningful content to users. Recommender Systems play a critical role in modern society and have transcended many different domains, pervading our lifes in advertising, e-learning, e-commerce, data analysis, online-dating, video game matchmaking, social networks, and many other domains.

These highly unique topics make generalizing Recommender Systems difficult, leading to many different solutions for many different but related problems. Combining some of these approaches, common techniques and findings regarding tasks such as user modeling or social recommendation become prevalent. Employing Recommender Systems in social environments, helping users to engage with the right person at the right time, becomes a possibility that could shape interactions between people in lots of domains.

While social contacts prove highly important in many aspects of life, social and cooperative learning are also known to positively affect learning outcomes. [1], [2] Successful group learning efforts can enhance cognitive and intellectual performance, student's social and communication skills and influence their overall satisfaction. [3], [4] Thus, social recommendation for learning is an important topic, helping to make mankind both more successfull and happier at the same time. **HIER VLL NOCH QUELLE AUS PAPER ZU LEUTE WERDEN BESSER BERUFSTAUGLICH.**

Besides the more traditional view of technology-enhanced learning, focussing on improving individual learning experiences by recommending exercises, media resources or additonal information at the right time depending on student's skills, preferences, needs and personality [5], [6], social learning research tried to connect learners to each other. Researchers explored opportunities for students to receive immediate peer support via online requests [7] or tasked students to discuss their answers to exercises in an e-learning environment with peers, which led to improvements in both short- and long-term performance. [8]

The goal of this report is to emphasize how Recommender Systems can be used not only to help a single learner solve problems or improve learning results, but to connect people, to build a community of learners and to enable students to engage in meaningful social learning opportunities. Along the way, relevant examples from other domains will be featured. After a quick introduction into the basics of user modeling and social recommendations in section II, section III will discuss a very recent paper regarding reciprocal recommender systems in learning environments. Finally, section IV will discuss the paper's proposed prototype and highlight some of the intricacies of reciprocal recommendation and learning group formation.

II. Related Work: Recommender Systems

The goal of a Recommender System [RS] is to emphasize relevant pieces of information in a convoluted stream of data, and to recommend a specific result to a specific user based on his or her history, preferences and situation. [9]

The following sections outline relevant findings and concepts regarding Recommender Systems in social environments and reciprocal peer recommendation. We will take a look at the

basics of user modeling in section, social recommendations and the importance of reciprocality.

A. User Modeling

In order to successfully recommend items, an RS needs to understand it's user. Goals, circumstances and domain-specific aspects need to be considered. Thus, RS need to model their users to try and understand what items might be relevant. Recommender Systems need to account for an adequate operationalization of the relevant personality traits, domain-specific preferences and surrounding circumstances and combine these into a user model.

Different domains or approaches within the same domain require different user models. For example, information about people connected to the current user can be helpful to improve recommendation quality. Based on user friendships and shared interests, specific items that one person liked could be recommended to friends [10] or just other users with comparable interests. [?] Another approach that was used to enable recommendations with scarce data and reduced computing power was to access a user's contacts and geographic history to find users with matching profiles and more information that can be used to improve recommendations. [11]

Even within a single domain, for example multiplayer videogame matchmaking, a system could focus on self-reported preferences [12], ratings by other players [13], implicit interaction-derived data [14], [15] or a combination of technical and self-reported information to improve the overall gaming experience. [16]

For example, recent study by Wang et al. [17] looked at the enjoyment of multiplayer gaming sessions in League of Legends (LoL) based on player personality. They followed a subset of Sternberg's [18] problem-solving styles in order to categorize the playstyle of different users. Wang et al. automated the data collection process with gameplay statistics. They assigned specific in-game actions to each problem solving style and determined a player's category based on his action profile. The results show a clear tendency of specific globally-active and risk-taking players to positively influence the overall game enjoyment, measured by the length of a game. It is important to note that the used measures are highly subjective to LoL.

In theory, user models want to encompass as much personalized information as possible, without becoming so convoluted that matching of items reverts back to being almost random. [19] Unfortunately, personalized data is hard to come by without having to ask the user directly, which can pose problems discussed in section IV.

For instance, self-reported information like a user's basic information, preferences or even personality could be considered to improve the user model. [20] This personality data could then be used in many different applications, from recommending jobs to movies for a group to watch. [21], [22]

As has already been established, different approaches fit different problems, although there is naturally some criticism conerning different types of data acquisition. Please refer to the sections III-A and IV for a deeper look into user modeling along a specific example.

B. Social Recommendation

In a world with an immense amount of possible social contacts, RS can help users to find other people to engage with, making recommendations relevant in social spaces. Lots of domain-specific social factors need to be considered in addition to the user model in applications in dating, learning, gaming, social networks or other domains.

In competitive multiplayer gaming, which is currently gaining in importance due to the increased interest in e-sports, the main goal is to create fair matches for two opposing teams. Usually, matchmaking in videogaming is concerned with bare player skills, but including the preferred style of playing, personality or character classes could lead to more balanced, fun games, making meaningful user models - again - highly important.

In the educational sector, research has focused on building meaningful professional engagements. A system could, for example, find a supervisor who fits a student's needs in competence, personality and topic expertise. [23] Or one could try to help researchers find meaningful partners on academic conferences based on shared study-interests and personality. [24]

On a lower level of competency, studying with peers is considered an especially effective way to improve lots of different skills and build knowledge. [4] When engaging in higher education, many students move to a different town and thus lack a social environment. This makes finding a study group a huge initial challenge. Considering the many theories concerning how effective learning groups should be structured (heterogenous with different skill-levels and a minimum joint skill, diverse in terms of gender and ethnicity, ... [2], [25]), finding an actually helpful group seems to be impossible. This opens another field of study: Peer recommendation in learning, which will be the main focus of this report. [19], [26]

Contrary to the aforementioned topics, where oftentimes a specific match for one user had to be found, group learning has to be beneficial for everybody involved. This adds another layer of complexity to this kind of recommendation: Reciprocality.

C. Reciprocal Recommendation

In extreme cases, Recommender Systems will have to recommend users to each other, forming reciprocal recommendations: A user receives other users as recommendations and is himself an item recommended to others. A true reciprocal recommendation is found when two users are recommended to each other.

This does not have to be the main goal: Systems trying

to recommend the best fit to each user without aiming for actual reciprocal recommendations could use simple scoring mechanisms. Each user receives scores describing their fit with other people. A certain amount of the highest scoring recommendations for each user will be returned. True reciprocal recommendations might happen as a byproduct of this process, but are not enforced or pushed, as explained in [26] and in section III-B2.

When advantages for all participating users are aspired, true reciprocal recommendations become a necessity. Special modeling techniques have to be employed to boost the recommendation strength of reciprocal recommendations and make these more likely.

For example, Xia et al. successfully designed a Reciprocal recommendation system for online dating, accessing self-reported user data and statistics of user's communication habits as a part of the network. [27] To determine recommendation scores, they used a similarity-based approach between users, incorporating:

- the user's general profile information
- the user's willingness to communicate with others
- the user's attractiveness to others, derived from how many other people contacted him or her

As an earlier paper revealed, these implicit, behavioral details proved to be much more relevant to actually predict user interactions than self-reported preferences in dating partners. [28]

The importance and applicability of reciprocal recommendations in learning scenarios will be highlighted using a recent study and will compose the main part of this report.
DAS IST WIEDER EIN TRANSITIONAL SATZ!

III. Reciprocal Peer Recommendation for Learning Purposes

With the goal of providing opportunities for meaningful learning engagements between learners, benefitting mutual success, Potts et al. introduce a novel algorithm and platform for reciprocal peer recommendation in learning environments. [26]

The scope of their report is to demonstrate the capabilities and explore the limitations of such a platform and algorithm on artificial data, before testing it under live conditions. Since the paper was published just recently before the writing of this report, further findings are not yet available. Lots of the theory on peer recommendation in social learning environments needs to undergo detailed testing and research. Some of the referenced papers are rather theoretical themselves, or offer insights in different domains and scenarios, making the transfer of knowledge difficult.

This chapter will cover the basics of *RiPPLE* as a peer recommendation platform and evaluational findings on artificial data. We will then discuss some shortcomings of the paper at hand and delve deeper into learnings from other studies and topics that might benefit the overall performance

of meaningful peer recommendation in the final section IV.

A. RiPPLE

RiPPLE ["Recommendation in Personalised Peer Learning Environments"] was designed and developed as a web-based online learning recommendation system. *RiPPLE* is an adaptive, student-facing, open-source platform with the aim to enable students to engage with others in meaningful learning experiences. To enhance the learning experience, *RiPPLE* functions as a learning platform, helping students to co-create and find meaningful learning-content and to find peers to learn with. This analysis will focus on *RiPPLE* as a peer recommendation platform.

Based on user input, *RiPPLE* will calculate potential matchups for its users. Depending on

- the competency derived by a user's performance on learning content
- his or her available timeslots
- the topics he or she would like to provide or seek peer support or find a learning partner in and
- the user's preferences on the respective skills of a potential partner in these topics

RiPPLE calculates a score for a matchup and will recommend a predefined amount of persons to each user as described in section II-C. As *RiPPLE* currently recommends learning opportunities for the upcoming week, updates to user preferences or competencies are represented once per week.

An important aspect of the recommendation algorithm is it's compatibility function, calculating a one-directional score for each combination of potential study partners, u_1 and u_2. In the first step, the algorithm will check whether a potential matchup is viable following two hard constraints:

1) a shared timeslot has to be available for both u_1 and u_2
2) the topic-specific joint competency must be greater than a prefedined threshold T. According to Blumenfeld [2], peer learning sessions will only become effective once the learners can share a minimum understanding of the topic.

For every pair of users satisfying these constraints, *RiPPLE* will calculate their respective one-directional scores. These represent how fitting u_2 is as a study partner for u_1 and vice-versa. (Since the users could have defined different preferences for their competency differences, scores don't need to be symmetric.) The scores take into account how good a matchup will be in terms of overall competency level, and how the other user matches the current users preferences. These values will be calculated across all topics relevant for u_1 and u_2. A visual representation of the resulting score can be seen in figure 1.

These two one-directional scores could now be used to find the the best partner for a specific user. To further recommend a matchup that is beneficial for both u_1 and u_2, the harmonic mean of both scores is considered as the "reciprocal score" of u_1 and u_2, a value that is now symmetric. [29] The
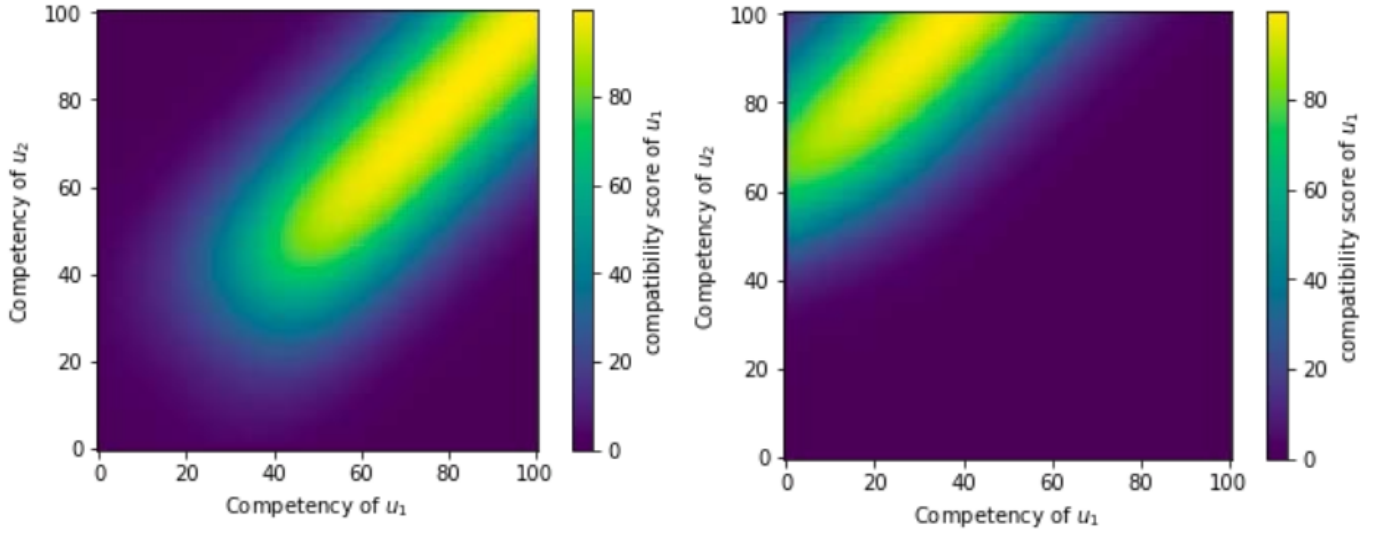
Fig. 1. The images show the areas of compatibility of a user u_2 as a function of u_1's competency. Lighter areas mean high compatibility scores in accordance to u_1's preferences. On the left u_1 is looking for a study partner, leading to the best fit along the $u_1 = u_2$ axis. The cutoff beneath a competency of 40 is due to the minimum joint competency threshold T. On the right, we can see u_1 looking for peer support, i.e. a person with considerably higher knowledge, here about 60 points higher than u_1. Source: [26]

harmonic mean, contrary to the arithmetic mean, pays respect to differences between it's values, making a larger gap between values less desirable. Peer-combinations with approximately similar scores will receive better final values, making matchups that are beneficial to both participants more relevant.

In the last step, *RiPPLE* returns a predefined amount of matchups k with the best reciprocal values for each user. Although these reciprocal values are now symmetrical, the recommendations don't have to be: While from u1s standpoint the matchup with u_2 and an (arbitrary) reciprocal score of 30 could be the very best opportunity, u_2 could still have true reciprocal a matchup with u_3 and a value of 50.

For more information on *RiPPLE*, the algorithm and further clarification of different variables, please reference [26].

B. Evaluation

In order to test *RiPPLE*'s applicability for actual use, Potts et al. designed an experimental setup in which *RiPPLE* would try to propose recommendations for randomly generated test data. Specific quality measures were designed according to [19] to assess different fields in which *RiPPLE* would have to show its capabilities. With satisfying results, *RiPPLE* would be able to be used under live conditions in the course of 2018.

To conduct the experimental evaluation, random data had to be generated; diverse enough to highlight edge cases but within reasonable bounds.

To fully satisfy as a tool recommending students to one another, *RiPPLE* must be able to form meaningful and successful matches for as many users as possible in reasonable time. On the other hand, minor drawbacks in the
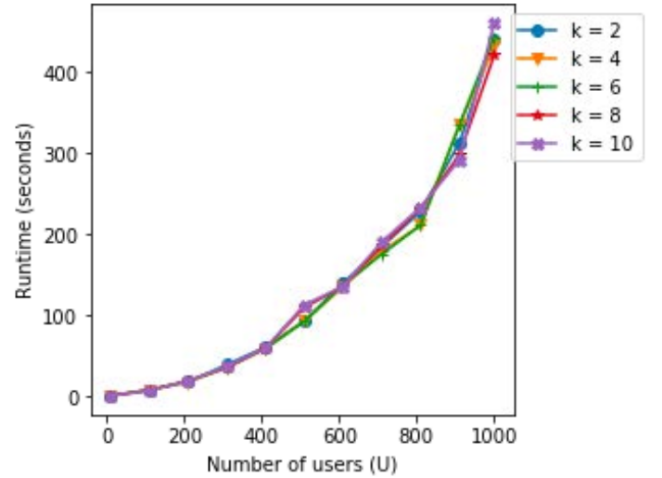

Fig. 2. Scalability: The algorithm's runtime depending on the number of users U and the amount of recommendations per user k. Note how k has almost no influence on the runtime, while it grows exponentially with increasing U. Source: [26]

defined metrics were considered to be tolerable in this step due to the experimental and randomly generated data and some further adjustments that could be made to compensate bad values.

As evaluation metrics for their experimental evaluation Potts et al. decided on four values that can further be used as general Quality Measures for reciprocal recommendation algorithms:
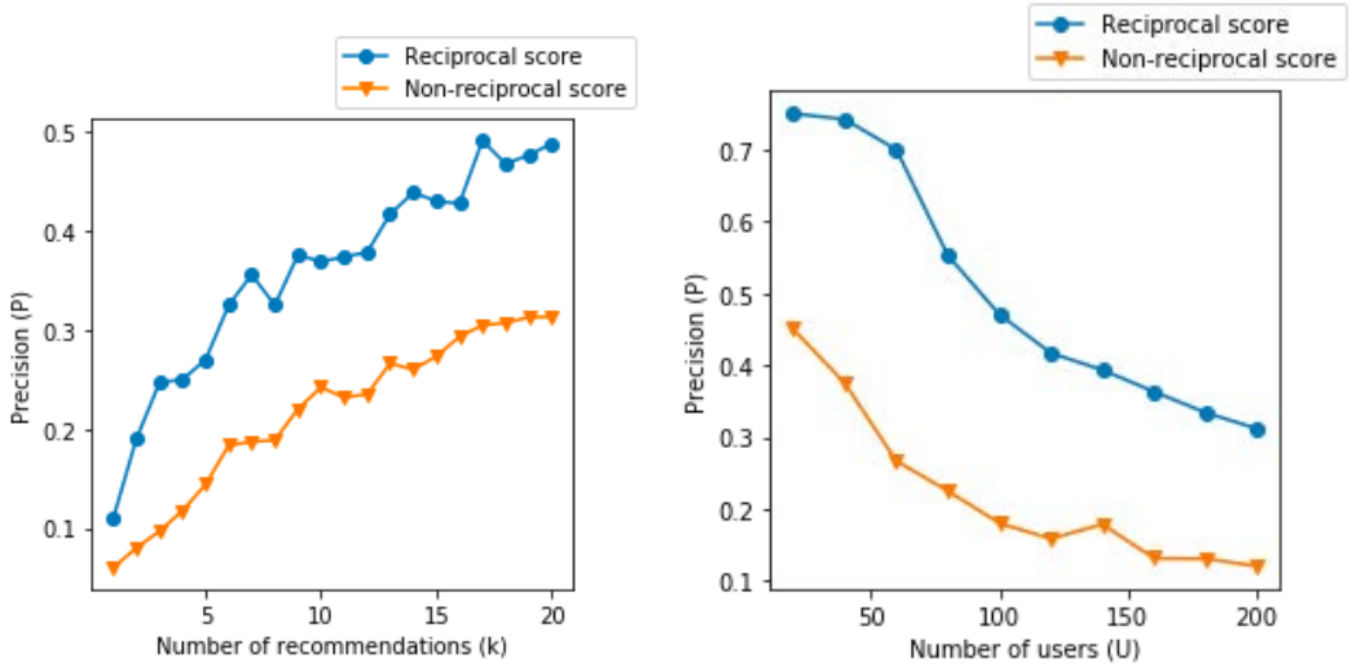
Fig. 3. Reciprocality: The precision (= the fraction of reciprocal recommendations out of the total recommendations averaged over all users) of the baseline non-reciprocal recommendations (orange) vs. of the reciprocal, averaged scores. Note how the reciprocal scores are always better. Source: [26]

1) Scalability: With increasing enrollment numbers in higher education, *RiPPLE* will have to be suitable for large sets of learners. High runtime and costs for evaluating datasets with reasonable amounts of students means slower responses and a worse user experience. An optimal solution could provide immediate recommendations to any user, at any moment.

As can be seen in figure 2, the runtime of the algorithm increased in a quadratic fashion, as U, the total amount of users, increased: $O(n^2)$. The number of recommendations per user, however, does not significantly impact the runtime. (Although the paper states that it *did* in fact affect runtime, looking at the plots data suggests that this might be a formatting error.)

Currently, *RiPPLE* calculates recommendations at the end of each week for the upcoming week, making the algorithm's runtime rather unimportant. In a 1000 user experiment, *RiPPLE* was able to provide recommendations for a single user in 0.045 seconds. However, further improvements are planned.

2) Reciprocality: The best possible recommendations are reciprocal: Users contacting a recommended user would also appear on this user's list of potential study partners. [29] Reciprocality was tested for both, the baseline non-reciprocal and the joint reciprocal harmonic mean scores. Whenever a user appears in the recommendations of a user on their own recommendation list that was built according to the respective score, the recommendation was considered to be reciprocal.

The precision for every user given the used score is calculated by dividing his reciprocal recommendations through k, the total amount of recommendations that user received. The system's total precision is defined as the average precision across all users. [29]

In all tested cases shown in figure 3, the reciprocal score had a higher precision than the baseline score. This is not surprising, since using the harmonic mean of both one-directional scores chooses reciprocal scores with medium values compared to non-reciprocal scores with a single high value. (As explained in section III-A) Increasing k also increases the precision, since more recommendations per user lead to a higher chance of reciprocal recommendations. On the other hand, increasing U with a fixed k reduces reciprocal precision, since there are more possible users to recommend.

3) Coverage: Recommending potential learning partners to one another should not abandone anyone. As such, coverage is a very important metric to consider. Since (almost) every user will receive recommendations, most users will be covered in one way or another. (The exception to this are users with completely incompatible timeslots, role preferences (i.e. being the only person looking for an equally skilled study partner) or users who can't meet the minimum competency when coupled with their available potential partners.) A very good fit can only be ensured when each user is recommended to others, ideally forming a reciprocal recommendation, which is represented in III-B2. Coverage however is defined as the percentage of users that appear in other's recommendations
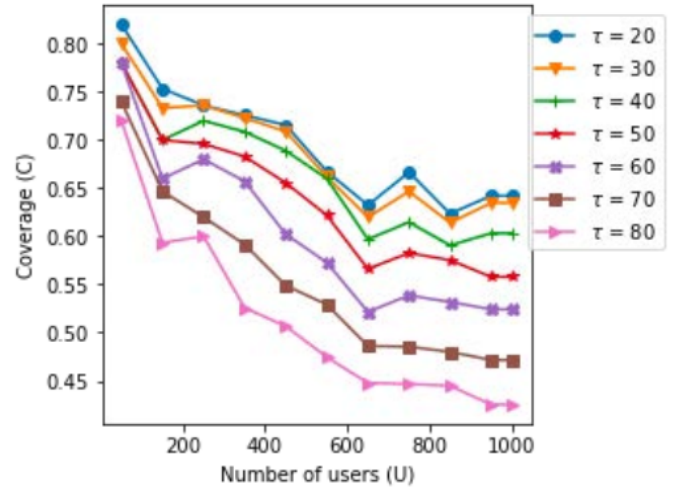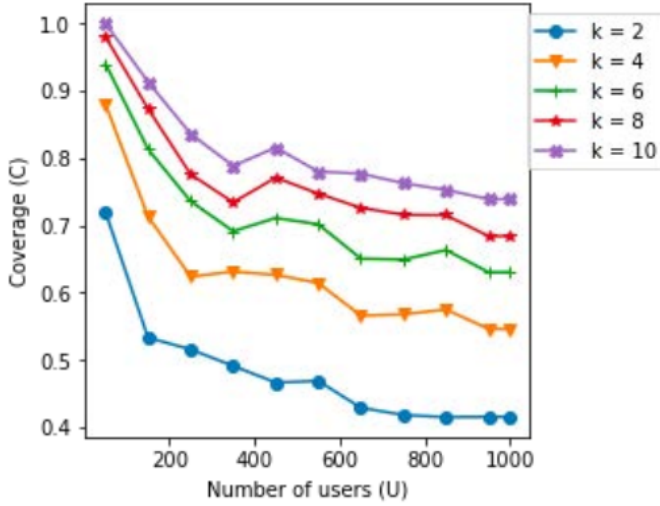
Fig. 4. Coverage: The percentage of users who appear in other's recommendations. With more users, coverage sinks (likelihood of hard to match users increases). Increasing received recommendations or lowering the minimum joint competency of matches increases coverage. Mind the y-axis cutoff. Source: [26]
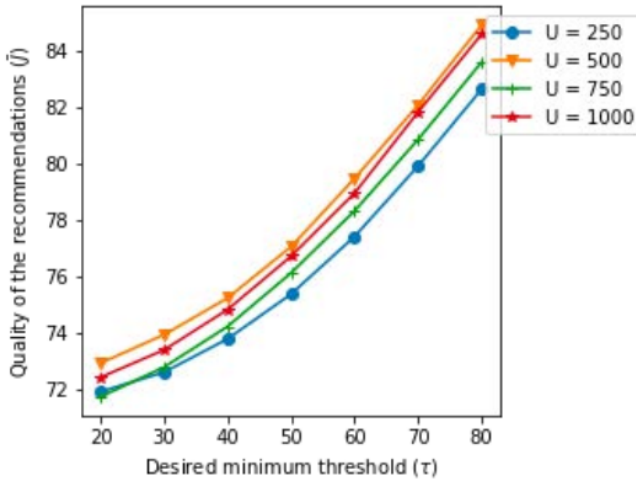

Fig. 5. Quality: The quality of matchups over different values for U and T. Note how changing U won't affect matchup quality. Choosing a higher minimum joint competency threshold τ for successful matchups increases overall quality. Note the different scaling of the axes, overemphasizing the quality increases. Source: [26]

at least once.

For a low amount of users and lots of recommendations per user, coverage is close to 0.9, meaning most users are recommended to others. As U increases or k decreases, the coverage sinks. However, more than 40% of users appear in other's recommendations under all tested circumstances. Reference figure 4 for a graphical overview.

4) Quality: The quality of a recommendation is not only based on its fit, but also on how good the resulting team

could perform. According to Blumenfeld, learners should meet a minimum competency level in order to be an effective group, as specified by the employed minimum matchup threshold τ. [2] Quality is thus defined as the user's average joint competencies across their matched topics. The goal is to generate matches that are as capable as possible in their respective fields of study.

As figure 5 shows, it is apparent that the total amount of users does not affect the quality of matches. The minimum threshold for joint competency of a matchup however leads to a better quality. Comparing this finding to figure 4 however, suggests that higher quality comes at the cost of less coverage. Especially when considering the slight improvements in quality score for larger increments in τ.

IV. DISCUSSION

Reciprocal peer recommendation is a highly promising topic with lots of applications, but is hard to handle, as a closer examination of "Reciprocal peer recommendation for Learning Purposes" and the implemented platform *RiPPLE* have shown. The system introduced by Potts et al. presents an approach to build a scalable, interactive and user-facing, multi-purpose platform to enhance learning both in on- and offline circumstances. While the experimental results look promising, further research on actual data has to be conducted, which is planned over the course of 2018.

The general concept shows some flaws as well.

The experimental evaluation of the platform's performance using artificial data sheds light on variable relationships, potential initial values for scientifically sparse concepts and the interplay of lots of factors. While the reported values

present good metrics to measure the algorithm's performance and suitability for live data, they don't actually evaluate the algorithm, since no target values have been set. Although the data is quantifiable, these findings should not be mistaken as quantitative results: instead of comparing the data to theory-driven goals and evaluating them for actual use, they are more or less providing an overview of how the algorithm works. In fact, there is currently no way to know if these results are "good". Some of the used measurements lack consistent data and research, and are not theoretically funded. For example, the question whether a coverage of little above 0.4 will be enough in practice, remains unanswered. Actual results from live usage are thus highly interesting and could provide insights in lots of different areas.

As a very recent system, *RiPPLE* was built to improve on earlier peer grouping algorithms.

A highly influential report by Olakanmi and Vassileva [19] highlights common shortcomings of peer recommendation studies and their proposed systems:

- Focus on improving learning instead of the goals of learning
- No information regarding the collection of user data
- Randomly assigning users in the first iteration and improving based on findings
- No consideration of scalability of the algorithm
- Inflexibility in working with fixed and limited constraints
- Too detailed user models leading to impossible matching
- Inflexibility in dealing with only partially known users
- Usage of self-reported user preference data
- Orphaned learners don't receive any value and have to be processed by hand
- No valid evaluation, just providing proofs of concept

RiPPLE was obviously built with these criticisms in mind. For example, RiPPLE aims to provide actual benefits to learners, uses a detailed user model and assigns students as informed as possible. Scalability and many other factors were considered when reporting the platform's applicability. *RiPPLE*'s flexibility in the definition of threshold values allows the tool to be customized to specific scenarios and needs.

However, many of the criticisms remain problematic until disproven: *RiPPLE*'s user model is light-weight but not evaluated and rather rigid, relying on the user's motivation to self-report data. The algorithm allows for learners to be orphaned due to different situations (as described below). Finally, as has already been stated, *RiPPLE* has not yet been evaluated in a real scenario.

Besides all this theoretical criticism, the algorithm itself does have some minor drawbacks. For one, *RiPPLE* calculates matchups across all topics. For example, two learners who would be a perfect match in one topic, but a bad match in another would be considered as a mediocre match. Topic-wise recommendation could further complicate the algorithm, but might lead to larger benefits for users.

Another way to improve the overall benefit to users would be to consider slightly larger groupts instead of matchups of two. Lots of theory about group composition emphasizes the importance of different skillsets and heterogeneity. [2], [19], [25] A system called DIANA, using a genetic framework to form small heterogenous groups in study courses used many different student characteristics to match students. An evaluation using course grades proved that this could generally be more efficient than random or self-organised group formation. Topic-wise recommendation would further complicate the algorithm, but might lead to larger benefits for users.**QUELLE**

The gaming sector could prove to yield some interesting findings regarding group formation, since lots of factors could be considered to build a well-performing multiplayer gaming team. [14], [15], [17] Unfortunately, this field of study is still sparse and inconsistent itself, facing struggles like different people choosing their preferred team based on completely different strategies. [12]

Another problem of *RiPPLE* are edge-cases in terms of competency preferences that would lead to overall low scores for matchups with other people, leading to learners who will receive suggestions with low scores, but won't appear in other's recommendations. While this does not necessarily lead to any consequences by itself, certain highly compatible users might be overwhelmed with meeting-requests from users outside their own recommendations. While they won't be able to meet every requesting user, these less compatible users might become abandoned.

Another drawback is the neglected human factor. Both user buy-in and competence in handling the tool and its demands might influence its use in practice. While this study's goal was explicitly to only test theory and future praxis tests are planned, this topic should be discussed, a major shortcoming of the paper at hand.

A lack in user buy-in is something that always should be considered, especially in a student context. If a student didn't want to engage with strangers, was not motivated to study with partners or to adjust his or her schedule, all recommendations to and of that student would not accomplish anything. Meeting requests would be ignored, and opportunities for matchups would expire. Even negative user manipulation needs to be considered as an possibility, but are something that has to be dealt with in the live test.

While missed opportunities are a problem of the students themselves, rather than of the platform providing recommendations, the other human factor needs to be addressed directly by the tool:

As humans are unreliable, self-reported metrics always underly lots of variance and errors. A user's competency in a specific topic, his or her preferences, or the willingness to commit to a specific timeslot for learning might change daily, depending on mood, time of day, culture and lots of other factors. [30] [31] Other variables, like a user's preferred skill difference towards a learning partner, are especially hard to specify. How is a user supposed to know what his or her learning preferences are? How would he know which

number refers to the desired difference in skill rating? From a psychological standpoint, this operationalization is bound to fail, if not controlled in an appropriate manner. [32]

This whole issue is not solely a problem of "Reciprocal peer recommendation for Learning Purposes". Generally speaking, choosing a user preference model that fits both the domain and the goal of an algorithm while still being able to perform in praxis poses a major problem in peer recommendation. [19], [26]

An important factor is to consider both, the information needed to create meaningful matches according to a specific success criterion, and how to access this information. Usually, automatically collected data is preferred to relieve the users as much as possible. But not all data can automatically be collected. Internal information, like preferences, date availability, motivations or the highly important factor of personality need to be reported by the user - there is currently no way to easily access internal information of the user's mind.

Psychological research was founded to enable the measurement and quantification of these details. From psychophysical measurements to intricate operationalizations of complicated internal states, one easily accessible method stays predominant in social science: self-reported statements, from qualitative interviews to acceptance scales.

One such scale is the NEO-PI-R [33], arguably one of the most famous personality-measurement tools. A questionnaire with Likert-scaled [1] items on five axes representing the five dimensions of personality. [35], [36]

While the NEO-PI-R is widely used and considered as reliable and validated, it still relies on self- or peer-reported data and can, as such, not be considered a flawless tool. For example, many self-reporting tools suffer from a relevant problem called "Faking Good", the tendency to answer items in a way that is considered to be socially acceptable. When faking good, people manipulate their answers to cohere to social norms out of fear to be seen as a bad person. "Faking Good" is known to influence the outcomes of some personality traits measured by the NEO-PI-R. [37] A similar effect can be observed when peer-ratings get influenced by peer sympathy. [38]

In learning environments, a phenomenon known as the "Lake Wobegon Effect" influences student reports of their learning success: students tend to overstate their good performances, while failures will not be reported. This leads to an overestimation of student successes in surveys. [39]

As another example, the infamous "Likert Scale" as introduced by Rensis Likert in 1932 as an "attitude scale" [34] can safely be assumed to be one of the most used metrics in social research, while it's optimal use still remains controversial. [30], [40]

Self-reported data stay a largely controversial topic in

psychological research: it is easy to acquire and enables researchers to access internal information, while these measurements can fluctuate following lots of different influences and are hard to validate. [30]–[32]

Not only is a proper specification of the domain and goals of a recommender system relevant, in reciprocal recommendation, the user must be modeled as both, a recommended item and a user receiving recommendations at the same time. Modeling users however is complicated due to unreliable methods or participants (knowingly or subconsciously) manipulating their answers.

While Potts et al. obviously tried to choose a user-model that is limited to the necessary basics and tried to rely on as few ambiguous and self-reported pieces of information as possible, they still need a user's ability in reporting information about him- or herself.

Other reciprocal recommendation approaches have to work around the same problems. As mentioned in section II-C, Xia et al. have shown that a behavior-emergent metric was more reliable in their use case of reciprocal online-dating recommendation. [28] Instead of focusing on user-reported data alone, they included implicit and pre-evaluated information derived from user interactions to measure attractiveness and willingness to communicate. [27] Wang et al. wanted to improve gaming matchmaking by employing problem-solving style information gathered from in-game statistics. They were able to deduce complicated, high-level cognitive problem-solving skills from simply evaluating implicit behavior, and used this information to improve player experiences. [17]

Instead of asking learners for their preferences regarding the competency difference towards their peer, Potts et al. could have decided the latter themselves, founded in theory of optimal group composition. [25]

Hier noch ein bisschen ins Detail zu gehen zu implicit/explicit

When talking about implicit data derived from a user's interaction with the running system on the other hand, an important drawback mentioned by Olakanmi et al. needs to be considered: the cold start problem. [19] Relying on live data will lead to the first recommendations to be made without any underlying data, at random. Only after some burn-in, actual data can be used to achieve better results, which leads to much less acceptance for the new tool.

As no satisfying approach to build effective groups in a new and untouched cohort of peers has yet been found, Potts et al. struggle to compile an effective and convincing user model themselves. Especially due to the many problems associated with measuring humans, algorithms are still unreliable when handling people in reciprocal situations.

Taking all of this into consideration, the peer learner recommendation algorithm proposed by Potts et al. does have its flaws but could theoretically benefit learners in praxis.

The applications of reciprocal peer recommendation are manifold and generally beneficial in a highly social and

---

[1]The commonly known items prompting users to reply to a statement on a scale from (usually) 1 to 5 (often combined with sentences like "I agree" or "I disagree") are called "Likert scales" after Rensis Likert. [34]

connected modern world. Bringing different people together and providing opportunities to engage with one another yields many advantages and teaches meaningful skills. However, human factors need to be incorporated in the algorithms to account for the dualistic role of users as both recommended items and recipients of recommendations. Many experiments with explicitly and implicitly collected data have uncovered a vast array of issues that need to be navigated. However, skillful design can circumvent some of these problems and create systems that are far from perfect, but nonetheless beneficial to their users. Time will tell if *RiPPLE's* streamlined user model and up-front evaluation of technical measures have made it a platform fit for practical use. Until then, further research is required.

References

[1] S. T. Bossert, D. C. Dwyer, B. Rowan, and G. V. Lee, "The instructional management role of the principal," *Educational administration quarterly*, vol. 18, no. 3, pp. 34–64, 1982.

[2] P. C. Blumenfeld, R. W. Marx, E. Soloway, and J. Krajcik, "Learning with peers: From small group cooperation to collaborative communities," *Educational researcher*, vol. 25, no. 8, pp. 37–39, 1996.

[3] C.-M. Zhao and G. D. Kuh, "Adding value: Learning communities and student engagement," *Research in higher education*, vol. 45, no. 2, pp. 115–138, 2004.

[4] J. Maxwell, "Learning together: Peer tutoring in higher education." *Journal of Physical Therapy Education*, vol. 22, no. 3, p. 97, 2008.

[5] H. Drachsler, K. Verbert, O. C. Santos, and N. Manouselis, "Panorama of recommender systems to support learning," in *Recommender systems handbook*. Springer, 2015, pp. 421–451.

[6] M. Erdt, A. Fernandez, and C. Rensing, "Evaluating recommender systems for technology enhanced learning: a quantitative survey," *IEEE Transactions on Learning Technologies*, vol. 8, no. 4, pp. 326–344, 2015.

[7] J. Greer, G. McCalla, J. Cooke, J. Collins, V. Kumar, A. Bishop, and J. Vassileva, "The intelligent helpdesk: Supporting peer-help in a university course," in *International Conference on Intelligent Tutoring Systems*. Springer, 1998, pp. 494–503.

[8] C. A. Reidsema, L. Kavanagh, E. Ollila, S. Otte, and J. E. McCredden, "Exploring the quality and effectiveness of online, focused peer discussions using the moocchat tool," in *27th Australasian Association for Engineering Education Conference*. AAEE, 2016.

[9] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender systems handbook*. Springer, 2011, pp. 1–35.

[10] H. Feng and X. Qian, "Recommendation via user's personality and social contextual," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 2013, pp. 1521–1524.

[11] L. Ramaswamy, P. Deepak, R. Polavarapu, K. Gunasekera, D. Garg, K. Visweswariah, and S. Kalyanaraman, "Caesar: A context-aware, social recommender system for low-end mobile devices," in *Mobile Data Management: Systems, Services and Middleware, 2009. MDM'09. Tenth International Conference on*. IEEE, 2009, pp. 338–347.

[12] J. Riegelsberger, S. Counts, S. D. Farnham, and B. C. Philips, "Personality matters: Incorporating detailed user attributes and preferences into the matchmaking process," in *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on*. IEEE, 2007, pp. 87–87.

[13] W. I. Patrick, S. D. Lamb, M. Bortnik, and J. P. Hansen, "System and method for providing feedback on game players and enhancing social matchmaking," 2011.

[14] M. Suznjevic, M. Matijasevic, and J. Konfic, "Application context based algorithm for player skill evaluation in moba games," in *Network and Systems Support for Games (NetGames), 2015 International Workshop on*. IEEE, 2015, pp. 1–6.

[15] O. Delalleau, E. Contal, E. Thibodeau-Laufer, R. C. Ferrari, Y. Bengio, and F. Zhang, "Beyond skill rating: Advanced matchmaking in ghost recon online," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 3, pp. 167–177, 2012.

[16] S. D. Farnham, B. C. Phillips, S. L. Tiernan, K. Steury, W. B. Fulton, and J. Riegelsberger, "Method for online game matchmaking using play style information," 2009.

[17] H. Wang, H.-T. Yang, and C.-T. Sun, "Thinking style and team competition game performance and enjoyment," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 3, pp. 243–254, 2015.

[18] R. J. Sternberg, *Thinking styles*. Cambridge University Press, 1999.

[19] O. A. Olakanmi and J. Vassileva, "Group matching for peer mentorship in small groups," in *CYTED-RITOS International Workshop on Groupware*. Springer, 2017, pp. 65–80.

[20] M. A. S. Nunes and R. Hu, "Personality-based recommender systems: an overview," in *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 2012, pp. 5–6.

[21] P. T. Costa Jr, R. R. McCrae, and G. G. Kay, "Persons, places, and personality: Career assessment using the revised neo personality inventory," *Journal of Career Assessment*, vol. 3, no. 2, pp. 123–139, 1995.

[22] J. A. Recio-Garcia, G. Jimenez-Diaz, A. A. Sanchez-Ruiz, and B. Diaz-Agudo, "Personality aware recommendations to groups," in *Proceedings of the third ACM conference on Recommender systems*. ACM, 2009, pp. 325–328.

[23] M. Zhang, J. Sun, J. Ma, T. Wu, and Z. Liu, "A personality matching-aided approach for supervisor recommendation (research-in-progress)," in *System Sciences (HICSS), 2016 49th Hawaii International Conference on*. IEEE, 2016, pp. 678–687.

[24] N. Y. Asabere, A. Acakpovi, and M. Michael, "Improving socially-aware recommendation accuracy through personality," *IEEE Transactions on Affective Computing*, 2017.

[25] S. Manske, T. Hecking, U. Hoppe, I.-A. Chounta, and S. Werneburg, "Using differences to make a difference: A study in heterogeneity of learning groups," in *11th International Conference on Computer Supported Collaborative Learning (CSCL 2015)*, 2015.

[26] B. Potts, H. Khosravi, C. Reidsema, A. Bakharia, M. Belonogoff, and M. Fleming, "Reciprocal peer recommendation for learning purposes," in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 2018.

[27] P. Xia, B. Liu, Y. Sun, and C. Chen, "Reciprocal recommendation system for online dating," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 2015, pp. 234–241.

[28] P. Xia, K. Tu, B. Ribeiro, H. Jiang, X. Wang, C. Chen, B. Liu, and D. Towsley, "Characterization of user online dating behavior and preference on a large online dating site," in *Social Network Analysis-Community Detection and Evolution*. Springer, 2014, pp. 193–217.

[29] S. Prabhakar, G. Spanakis, and O. Zaiane, "Reciprocal recommender system for learners in massive open online courses (moocs)," in *International Conference on Web-Based Learning*. Springer, 2017, pp. 157–167.

[30] J. W. Lee, P. S. Jones, Y. Mineyama, and X. E. Zhang, "Cultural differences in responses to a likert scale," *Research in nursing & health*, vol. 25, no. 4, pp. 295–306, 2002.

[31] J. Sorensen, "Measuring emotions in a consumer decision-making contextapproaching or avoiding," *Aalborg University, Denmark*, 2008.

[32] R. M. Gonyea, "Self-reported data in institutional research: Review and recommendations," *New directions for institutional research*, vol. 2005, no. 127, pp. 73–89, 2005.

[33] F. Ostendorf and A. Angleitner, *NEO-Persönlichkeitsinventar nach Costa und McCrae, Revidierte Fassung (NEO-PI-R)*. Hogrefe Göttingen, 2004.

[34] R. Likert, "A technique for the measurement of attitudes." *Archives of psychology*, 1932.

[35] R. R. McCrae and P. T. Costa, "Validation of the five-factor model of personality across instruments and observers." *Journal of personality and social psychology*, vol. 52, no. 1, p. 81, 1987.

[36] L. R. Goldberg, "An alternative "description of personality": the big-five factor structure." *Journal of personality and social psychology*, vol. 59, no. 6, p. 1216, 1990.

[37] B. Griffin, B. Hesketh, and D. Grayson, "Applicants faking good: evidence of item bias in the neo pi-r," *Personality and Individual Differences*, vol. 36, no. 7, pp. 1545–1558, 2004.

[38] D. Leising, J. Erbs, and U. Fritz, "The letter of recommendation effect in informant ratings of personality." *Journal of Personality and Social Psychology*, vol. 98, no. 4, p. 668, 2010.

[39] N. L. Maxwell and J. S. Lopus, "The lake wobegon effect in student self-reported data," *The American Economic Review*, vol. 84, no. 2, pp. 201–205, 1994.

[40] L. Chang, "A psychometric evaluation of 4-point and 6-point likert-type scales in relation to reliability and validity," *Applied psychological measurement*, vol. 18, no. 3, pp. 205–215, 1994.