Team Dating: A Self-Organized Team Formation Strategy for Collaborative Crowdsourcing

Ioanna Lykourentzou
Luxembourg Institute of
Science and Technology
Esch-sur-Alzette, Luxembourg
and
Carnegie Mellon University
Pittsburgh, PA 15213, USA
ioanna.lykourentzou@list.lu
ilykoure@andrew.cmu.edu

Robert E. Kraut
Carnegie Mellon University
Pittsburgh, PA 15213, USA
robert.kraut@cs.cmu.edu

Shannon Wang
Carnegie Mellon University
Pittsburgh, PA 15213, USA
xueyingw@andrew.cmu.edu

Steven P. Dow
UC San Diego
San Diego, CA 92093, USA
spdow@ucsd.edu

Abstract
Online crowds have the potential to do more complex work in teams, rather than as individuals. However, at such a large scale, team formation can be difficult to coordinate. (How) can we rely on the crowd itself to organize into effective teams? Our research explores a strategy for "team dating", a self-organized crowd team formation approach where workers try out and rate different candidate partners. In two online experiments, we find that team dating affects the way that people select partners and how they evaluate them. We use these results to draw useful conclusions for the future of team dating and its implications for collaborative crowdsourcing.

Author Keywords
Collaborative crowdsourcing; team building; self-organization.

ACM Classification Keywords
H.5.3 [Group and Organization Interfaces]: Collaborative computing; Computer-supported cooperative work

Introduction
Collaborative crowdsourcing, i.e. the type of crowdsourcing that relies on teamwork, is often used for tasks like product design, idea brainstorming or knowledge development. Although the effective team formation may enable crowds

Is there a task-independent, scalable and on-the-fly way of building crowd teams that captures both skills and interpersonal dynamics?

Team dating: Delegating team building to the crowd workers themselves.

A try-and-evaluate process: People try out different co-workers, discuss with them, and indicate those that they prefer working with.

It functions in face-to-face teams: Short initial interactions between pairs of students have helped gather interpersonal evaluations and create better teams in traditional classroom settings.

How will crowd workers react to this new team building process?

Figure 1: Team dating key points

to do more complex and creative work, forming teams in a way that optimizes outcomes is a new area for research.

In traditional settings, like corporations [5] or inside a classroom [2], team selection benefits from referrals, the knowledge of people's skills and preferences that a good manager, teacher or member of the HR may have, and on the personal opinion of the latter as to whether certain people can work together effectively [4]. In crowdsourcing however this is often difficult, due to the scale and lack of data on task-specific worker history, especially for collaborative tasks. Most importantly, due to the lack of a shared work identity and culture (as it is the case in corporations), crowdsourcing environments make it very difficult to estimate the internal dynamics and interpersonal relationships between the members of a future crowd team. These internal relationships, like providing and accepting feedback, cooperation and communication quality, team spirit and morale, adaptability and coordination [5], are nevertheless extremely important for effective team formation.

This research explores how to dynamically create crowd teams from a large population of potential workers without any prior knowledge of worker profiles. We propose a method that can capture both the skills and the interpersonal work dynamics of the specific workers for the specific task: team dating. The idea behind team dating is simple: what if we delegate team building to the crowd workers themselves? What if we ask them to try out different candidate co-workers, evaluate them, indicate those that they like working with, and make crowd teams based on these indications?

Team dating is based on decentralized, as opposed to centrally-coordinated, team building. Decentralization as a means of decision-making has proven beneficial in many different settings, ranging from multi-agent systems and distributed artificial intelligence [7] to peer evaluation in education [3]. For example, in the same sense that peer assessment can help distribute the time and effort required to evaluate a set of student essays, team dating may help distribute the time and effort needed to evaluate the skills of a set of crowd workers. Moreover, team dating could be an excellent way of evaluating the interpersonal relations between the specific workers and for the specific task, something that is hardly achievable by evaluation methods targeted at individual workers.

Overall, team dating could potentially help improve crowd team building in three aspects:

- Filter out non-performing workers.
- Provide tailored, on-the-fly and task-specific skill and interpersonal qualities estimation.
- Eventually help build better teams.

In this late-breaking work we present the team dating idea (Figure 1) and a method for implementing it in practice. We also examine a set of questions related to the way that crowd workers function inside a team dating environment: how do people evaluate their candidate co-workers? How do they choose their preferred teammates? How do their evaluations compare to external evaluations from people that have not interacted with them? These preliminary results are meant to examine whether team dating is worth investigating further or whether, despite its optimistic assumptions, this method would be similar to simply applying an automated profiling step prior to the main collaborative crowdsourcing task.

The rest of this paper is organized as follows: first we explain the team dating process. Then we present our preliminary results, focusing on the pre-team building phase, i.e. on the effect that team dating has on worker selections and

Figure 2: Overview of the team dating process.

its relation to external evaluations. We close with a discussion of the future directions of this work and its implications for collaborative crowdsourcing and HCI research.

Team Dating Overview

The process in a nutshell
The team dating process (Figure 2) works as follows: first we hire a worker batch[1] and ask each worker to perform a short individual task, related to the main task that the final teams will work on. The task used in this paper is slogan creation for a coffee advertisement (detailed in the following). Next, we start the team dating phase, in which we pair each worker with a randomly selected co-worker, for an X number of times (X being the number of team dating rounds). As soon as the team dating phase is over, we ask all workers to indicate which colleagues they would like to work with. Their selections are to be used for creating the final teams. In more detail:

Preparation phase: Worker Batch hiring
First we hire a batch of workers. The workers arrive to a "preparation" web page, where they fill in some basic demographic information (age, gender, educational level, ethnicity) and they are explained the team dating process. This page is also meant to keep workers busy while a sufficient worker number has been gathered, and the process can start.

The team dating process heavily relies on synchronization: from the next step onwards all workers proceed at the same pace, being redirected at the same time from page to page, to ensure that they will all have a team dating partner at each round.

---

[1] In compliance with the technical requirements of our team dating application software, workers are hired in batches to ensure a seamless experience of synchronous interaction.

Phase 1: Individual task - Make a slogan
In this step workers are asked to make a slogan for the advertisement of a product. The product used in this work is a fictive coffee beverage called sCOPA, and its properties are communicated to the workers through a description that is short and easily understandable, but at the same time gives enough details to stimulate the creation of diverse slogans. It is also meant to describe an easily recognizable product (coffee) but without reminding any specific brand (like a cola product would). Workers are asked to make a slogan targeted for TV broadcasting of the above product. They are advised that their product should be original, understandable and emphasize the unique aspects of the advertised coffee beverage.

The purpose of this step is to familiarize the workers with what they will eventually have to work on, so that they will respectively be able to evaluate their peers' skills on the specific task. The individual slogans are also going to be given as a starting point of conversation between the team dating worker pairs in the next phase.

Phase 2: Team dating - Try out your peers
In the team dating phase, each worker is paired with a co-worker, randomly chosen from the batch. The two workers enter a dedicated chat room and discuss for a short period of time (3 minutes in our experiment as detailed below). During this time, the workers are asked to read each other's slogan, and try to discover if they would like to work together. Figure 3 illustrates what two example workers, Anna and Peter would see in this phase.

In this exploratory study, the team dating discussion topic is left open for the workers to decide. This choice has been made for two reasons. First, leaving the topic open seems to enable people to evaluate each other's skills and interpersonal work relations, as reported by one of the few stud-
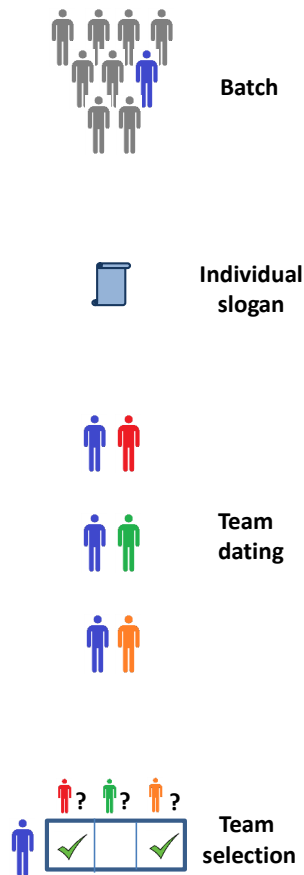
Hi **Anna** and **Peter**!

You have been paired to chat together for 3 minutes. Your goal is discuss and figure out if you would like to work together inside a team in the next phase (full advertisement). When time is up you will be prompted to evaluate each other as a potential teammate.

Hint: To see if you can work well together, you can try to make a common slogan from your two individual slogans.

Anna's slogan: **A cup, a smile, a coffee sCOPA for you**

Peter's slogan: **Green coffee for a clean living**

Figure 3: In each team dating round two random workers, here Anna and Peter, are asked to discuss to see if they can work together.

ies using this method, albeit for face-to-face interactions and educational contexts [2]. Second, choosing not to direct the conversation towards a specific topic seems appropriate so as to not confound the results of this exploratory study, but rather to allow all possible discussion patterns emerge and possibly help us refine future implementations of the process. Keeping in mind however that people may feel reluctant contributing totally from scratch (a phenomenon observed in other cases of collaboration like wikis [6]), we do provide them with each other's slogans and an open-ended hint of combining them, to serve as the discussion starting point. As soon as the discussion time is up, both workers are simultaneously asked to rate each other. In our example scenario, Figure 4 illustrates what Anna sees.

Both questions are rated using a 5-item Likert scale from "Not good/well at all" to "Excellent". After each team dating and evaluation round, each person is paired with another randomly selected co-worker and the process is repeated.

Phase 3: Teammate selection
After the team dating phase is over, we ask each worker to indicate the people he/she would like to work with in the main task, by providing them with a list of all workers of the batch. Multiple pieces of information, gathered from the previous two phases can be given for each candidate co-worker: i) *Demographic data*, ii) the *individual slogan* of the worker, iii) the *personal rating*, in case the rater and

1. How good will Peter be in the main task?

2. How well do you think you will be able to collaborate with Peter as a team in the main task?

Figure 4: After talking to Peter during their team dating session, Anna is asked to rate him on skill and compatibility.

the ratee have discussed during team dating phase and iv) *collective rating*, i.e. the average rating of the worker from all the people she/he worked with during team dating. After making their choice of preferred teammates, the workers are moved to the main task, where they will work on an extension of the individual task, i.e. on making an advertisement script of the same product (full advertisement). The exact way of building the teams and the configuration of the main task phase are left for a subsequent study, since in this late-breaking-work we are interested specifically in the way that team dating affects worker selections rather than their final main task outcome.

Preliminary Results

Experimental setup
For this experiment we hired two worker batches. Batch 1 consisted of 27 workers, who were given access to all four information parameters in the teammate selection phase (demographics, individual slogans, individual ratings and collective ratings). Batch 2 consisted of 29 workers who decided only on the basis of collective ratings. All workers were recruited through the CrowdFlower platform, they had not worked together before on the task and they were paid $4 per person. Workers spent three minutes on the individual task of Phase 1. In Phase 2 they participated in three team dating rounds of four minutes each (three minutes discussing and one minute in evaluating their co-worker). Finally they had five minutes to select their teammates in Phase 3. These timings were chosen after prior piloting to determine the minimum amount of time that allows meaningful interactions among the workers for the specific task. The total duration of the experiment was 20 minutes. Results were as follows (Figure 5):

Team dating affects the way people evaluate their co-workers: Judgments after participating in the team dating differ than judgments by external reviewers.

Workers trust their personal judgments in selecting teammates: In selecting who they want to work with inside a team, people trusted their own rather than the average ratings.

Figure 5: Results in a nutshell

How do workers evaluate candidate co-workers?
First we examine the way that the workers evaluate their co-workers at the end of the team dating phase. We remark two observations:

- Workers are rated consistently by different teammates.

Each worker is rated by three co-workers (one per speed dating round). Their group ratings give an intra-class correlation value ICC=.61 for both batches, which signifies that different raters agree to a reasonable extent in their evaluations regarding the rated workers.

- Evaluating for skill and compatibility basically mean the same thing for the workers.

In both batches, the evaluations the workers gave as to the skill and the compatibility of their evaluated co-workers were very similar (rs=.86 in Batch 1 and .97 in Batch 2).

How do workers choose their teammates?
Next we examine the way that workers decide on their preferred teammates, at the end of phase 2.

- Having worked with someone during team dating does not necessarily mean selecting them for the main task.

A random-effects logistic regression analysis using the Wald Chi-Square Test shows that for Batch 2 (where workers saw only the average quality judgments when selecting preferred teammates), both working with a person and the person's average quality increases the odds of selecting them for the main task, 6.4 and 2.2 times respectively (with z=5.17, p<0.001 for working with the person and z=3.72, p<0.001 for average quality). This however is not true for

Batch 1, where workers had four inputs for deciding preferred teammates. In this case, neither working with a person nor the person's average quality score predicted selecting them for the main task (OR=1.64, z=0.76, p=0.45 for the working with the person coefficient and OR=1.02, z=0.1, p=0.92 for the average quality one).

- People select teammates based on their personal opinion, not influenced by others.

Performing a similar analysis as above, we observe that the individual but not the collective judgments predict the selection of people to work with in the main task. Specifically, the individual judgments increase the odds of selecting a person 3.5×10^6 times for Batch 1 and 26.63 for Batch 2 (z=2.01, p=0.045 and z=1.80, p=0.072 respectively), while one standard deviation increase in the average quality increases the odds of selecting them 1.23 times for Batch 1 and 6.74×10^{-6} for Batch 2 (z=0.23, p=0.82 and z=-1.32, p=0.19 respectively).

How does team dating compare to external evaluations?
Despite its advantage in gathering task-specific evaluations, team dating is a costly procedure. A natural question we therefore need to answer before closing this study is: "Could we get the same evaluations without the team dating phase?" To answer this, we evaluated the individual slogans of each worker using 10 external evaluators, and averaged their judgments. The external evaluators rated the slogans on 6 quality axes drawn from content quality literature [1]: i) Originality, ii) Honesty, iii) Customer attraction potential, iv) Simplicity, v) Uniqueness and vi) Overall Impression. Results showed that the external judgments are not highly correlated with the judgments after team dating (rs=.18 in Batch 1 and .03 in Batch 2, on the "Overall Impression" quality axis. Similar results for the other quality axes.) This implies that the team daters actually use their

interaction with their partner to evaluate them, and they do not rely only on their partner's "objective skill".

Discussion and Further work

A number of useful remarks can be drawn from the above-presented experiments, to guide further work on team dating. The fact that people are rated consistently by their peers indicates that the method could potentially be trusted to provide on-the-fly evaluations of specific workers hired for a specific task. In parallel, the fact that skill evaluations after team dating are different than skill evaluations based only slogans only implies that the team dating approach changes the manner that a person's skills are perceived and eventually rated. It remains to be seen whether these evaluations can be combined, and how, for forming more successful crowd teams.

In terms of future improvements, the above results indicate two issues. First, the fact that skill and compatibility estimation are practically indistinguishable for most workers means that the two questions could potentially be merged into one summarizing the person's opinion of their team date. It could however also mean that more refined questions are needed to extract useful judgments from the workers, to guide an effective subsequent team building process. Second, the fact that personal judgments dominate group (collective) judgments in selecting future teammates reduces the need for a decision page, and indicates that people's opinions from the team dating round could be used directly to form the main task teams.

Finally in terms of future extensions, a number of interesting directions can be foreseen. First, a qualitative study on the content of worker discussions could give useful insights regarding the nature of worker interactions, and help us design a more effective team dating process. Given the time pressure that team dating entails, it is also of particular interest to explore whether one should guide worker discussions more explicitly, and which guidance method allows the workers to discover each other's skills and work styles more efficiently. Last, a comparative study, applying team dating on crowd versus face-to-face teams (an extension of the study proposed in [2]), could be also useful to reveal the relation between technology and the factors that govern decision-making of teammate selection.

Acknowledgements

REFERENCES

1. Kevin Chai, Vidyasagar Potdar, and Tharam Dillon. Content Quality Assessment Related Frameworks for Social Media. In *Proceedings of the International Conference on Computational Science and Its Applications: Part II*. Springer-Verlag, Berlin, Heidelberg, (2009), 791–805.

2. Petru L. Curseu, Patrick Kenis, Jorg Raab, and Ulrik Brandes. Composing Effective Teams through Team Dating. *Organization Studies* 31, 7 (2010), 873–894.

3. Nancy Falchikov and Judy Goldfinch. Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research* 70, 3 (2000), 287–322.

4. Daniel P. Forbes, Patricia S. Borchert, Mary E. Zellmer-Bruhn, and Harry J. Sapienza. Entrepreneurial Team Formation: An Exploration of New Member Addition. *Entrepreneurship Theory and Practice* 30, 2 (2006), 225–248.

5. John R. Hollenbeck, D. Scott DeRue, and Rick Guzzo. Bridging the gap between I/O research and HR practice: Improving team composition, team training, and team task design. *Human Resource Management* 43, 4 (2004), 353–366.

6. Ioanna Lykourentzou, Younes Djaghloul, Katerina Papadaki, Foteini Dagka, and Thibaud Latour. Planning for a Successful Corporate Wiki. In *Communications in Computer and Information Science, Digital Enterprise and Information Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg (2011) 425–439.

7. Gerhard Weiss. *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press, Cambridge Massachusetts (1999).