Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media

Crowdphysics: Planned and Opportunistic Crowdsourcing for Physical Tasks

Adam Sadilek[*]
Department of Computer Science
University of Rochester
Rochester, NY 14627
sadilek@cs.rochester.edu

John Krumm
Microsoft Research
One Microsoft Way
Redmond, WA 98052
jckrumm@microsoft.com

Eric Horvitz
Microsoft Research
One Microsoft Way
Redmond, WA 98052
horvitz@microsoft.com

Abstract

Research on human computation and crowdsourcing has concentrated on tasks that can be accomplished remotely over the Internet. We introduce a general class of problems we call *crowdphysics* (CP)—crowdsourcing tasks that require people to collaborate and synchronize *both in time and physical space*. As an illustrative example, we focus on a crowd-powered delivery service—a specific CP instance where people go about their daily lives, but have the opportunity to carry packages to be delivered to specific locations or individuals. Each package is handed off from person to person based on overlaps in time and space until it is delivered. We formulate CP tasks by reduction to a graph-planning problem, and analyze the performance using a large sample of geotagged tweets as a proxy for people's location. We show that packages can be delivered with remarkable speed and coverage. These results hold for the case when we know people's future locations and also when routing without global knowledge, making only local greedy decisions. To our knowledge, this is the first empirical evidence that dynamic networks of mobile individuals are highly navigable.

Figure 1: A package on its way to a destination in Seattle, routed among geolocated Twitter users (colored pins).

Introduction

Can we effectively deliver packages with the crowd? We seek in this paper to answer this concrete question, and more generally to explore the potential and limits of a broad class of crowdsourcing problems that require people to synchronize actions in time and physical space.

Research on crowdsourcing has been evolving on multiple fronts. Studies have demonstrated that the crowd can sometimes be harnessed to solve tasks more effectively and accurately than a single expert. Successes with leveraging the crowd has influenced thinking within a wide range of disciplines, from psychology to machine learning, and includes work on crowdsourcing diverse tasks such as text editing (Bernstein et al. 2010), image labeling (Von Ahn and Dabbish 2004), speech transcription (Lasecki et al. 2012), language translation (Shahaf and Horvitz 2010), software development (Little and Miller 2006), and providing new forms of accessibility for the disabled (Bigham et al. 2010).

Most work to date has concentrated on the solution of intellectual challenges that can be accomplished in their entirety remotely over the Internet. Moreover, workers have been engaged largely to participate in tasks independently of one another. We argue that computer-aware connectedness among people has reached an inflection point that supports a paradigm shift in thinking about coordination and collaboration on physical activities. With this framing, we explore the crowdsourcing of tasks that require people to sequence or synchronize physical actions in time and space. We call this broader class of problems *crowdphysics* (CP). CP solutions harness programmatic access to people like other crowdsourcing, but draw upon sensing, inference, and prediction to guide action in the physical world. We see great opportunities to leverage methods and results from graph theory, optimization, and machine learning to solve CP challenges.

We introduce and explore a canonical example of CP with a prototype delivery system we call *TwedEx*. We assume that people go about their daily lives, but have the opportunity to carry a package (*e.g.*, mail) to be delivered to specific locations or individuals. Each package is handed off from person to person based on overlaps in time and space until a target location or person is reached.

[*] Adam Sadilek performed this work at Microsoft Research.

Numerous tasks can be solved via harnessing synchronized efforts of a dynamic mesh of people on the move. Instances include building a sensor network of people to assist in finding a missing child and summoning a team to a location to help retrieve a lost object (*e.g.*, a wallet). On a larger-scale, CP systems could assist with such tasks as coordinating humanitarian efforts in the developing world (*e.g.*, distributing supplies or vaccines), providing search and rescue in times of crisis (*e.g.*, assisting people affected by a natural disaster or amidst war), and protecting citizens from an oppressive regime. Such efforts may leverage social connections of different degrees per preferences about trust. However, they can also be constructed from strangers, who may not even be aware of the other parties in a coordinated task.

Another task from the CP class is disease containment. Preventing an outbreak of an infectious disease is analogous to package delivery with the cost function inverted. The pathogen becomes the package,[1] the contact network is induced from all potential package handoffs, and we now want to *minimize* the probability of effective spread. Coordinated CP tasks may be used to fight the spread of disease by motivating strategically selected people to change their patterns of mobility, including changing the portion of time they stay at home, avoiding global travel of certain types, making substitutions in destinations, and avoiding specific venues. Prior work explored disease containment, but only at an aggregate level and with simulated populations (Eubank et al. 2004), but the increased level of detail of online data now enables us to capture specific individuals at a population scale (Sadilek, Kautz, and Silenzio 2012).

Our intuition is that if package delivery can be solved, numerous other CP problems can be *reduced* to it. On the specific example of package delivery, we have found that the CP approach can actually be competitive with traditional delivery services in several ways. However, we present the delivery task primarily as a motivating example for research on physical crowdsourcing and intelligent coordination of the crowd.

A successful delivery requires a chain of correct handoffs between pairs of workers. Each exchange requires the participants to meet (*i.e.*, to be synchronized in time and space). There needs to be global synchrony *across* handoffs to create the desired delivery path. The summon task is simpler: a set of workers only have to arrive at a preset constant location. A delivery task can be *decomposed* into a sequence of synchronized summon tasks—one for each handoff, summoning two workers—to locations that minimize a given cost function. This function can trade off quantities such as task cost, digression of workers from their intended paths or routines, reliability, and speed of task completion.

Significance of Results

Focusing on the crowdsourcing of package delivery as a specific instance of CP, we show that the physical system has permeability, which increases considerably as we allow longer wait times and larger geographic digressions of workers in defining feasible package handoffs. We explore a novel setting of delivery to a moving person and find that over 40% of all pairs of people are reachable within ten days in Seattle alone. We distinguish between idealized global routing for optimal handoffs, where we assume we know where and when everyone will be in the future, and more realistic routing under uncertainty, exploiting local routing. In the latter setting, we assume we know basic statistics about people's past locations. We show that local routing algorithms working under uncertainty do not significantly degrade performance as compared to an ideal global optimum. Finally, we demonstrate that efficient global delivery is possible, with delivery times often comparable to direct travel times among locations.

[1] Unlike a package, the pathogen can "divide" and infect multiple individuals at once.

Dataset	Days	Users	Tweets	Edges in G
NYC	21	47,713	544,606	740,489
SEA	21	10,424	125,620	140,075
US	2	371,481	3,434,898	3,931,884

Table 1: Summary statistics of the three datasets—New York City (NYC), Seattle (SEA), and continental United States (US)—used in experiments below after filtering out duplicate and questionable tweets. Number of edges in routing graph G is for following parameter settings: $\delta = 100$ meters and $\tau = 0.5$ hours (see Methodology section for definition of G).

Dataset

To evaluate our approach on large-scale real-world data, we leverage a sample of GPS-tagged tweets originating from the United States over a period of 21 days (starting on March 1, 2012). The experiments below concentrate on three geographical areas that are subsets of the larger dataset: New York City (NYC), Seattle (SEA), and the continental US. For NYC and SEA, we use tweets from within a 60×60 kilometer bounding box centered over each city.

Since we explore whether real individuals in a large population can deliver packages collectively, we eliminate tweets that are clearly anomalous or come from multiple accounts (per removing likely robot-generated spam). Specifically, we eliminate users that produce tweets too quickly and too far apart given typical driving and flying speeds. Out of the remaining users, we keep only those who tweeted at least twice in the data collection period, as people with only one data point cannot participate in package delivery. Table 1 summarizes three subsets of the dataset used in the experiments. The statistics are computed *after* the above filtering has been applied.

In agreement with previous work, we find that both the length and duration of "flights" of Twitter users (*i.e.*, two consecutive tweets of a user) exhibit a consistent power-law distribution in the form $P(x) \propto x^{-\beta}$ with $\beta \approx 2.2$. This type of distribution with β ranging from 1.5 to 3 arises in many natural systems that involve human activities, including the behavior of cell phone subscribers and epackage users (Wu et al. 2004; González, Hidalgo, and Barabási 2008), academic citations (Leskovec, Kleinberg, and Faloutsos 2005), dispersal of bank notes (Brockmann, Hufnagel, and Geisel 2006), Internet connectivity (Faloutsos, Faloutsos, and Faloutsos 1999), Web link structure (Barabási and Albert 1999), Einstein's correspondence patterns (Oliveira and Barabási 2005), and a wide array of phenomena in online social networks (Leskovec and Horvitz 2008; Cheng et al. 2011).

Our observed distribution indicates that geotagged Twitter activity can be accurately modeled at a high level as a Lévy flight random walk characterized by a mixture of frequent short displacements and relatively rare but long jumps (Mandelbrot 1982). By contrast, networks that do not evolve for locally navigability tend to have a significantly different β. For instance, the node degree distribution of the US power grid has $\beta \approx 4$ (Barabási and Albert 1999).

The networks commonly seen in natural systems involving human activity often exhibit the "small world" phenomenon, where the diameter is $O(\log n)$ for networks with n nodes (Milgram 1967). This means that even large networks contain very short (exponentially shorter) paths among pairs of nodes. For the package delivery application, the question is whether these short paths can be found and leveraged by individual nodes in a *dynamic* network when operating with only local knowledge. As we will see, this frequently can be done.

Prior theoretical work has concentrated on quantifying the navigability of static homogeneous networks with a clean, repeated lattice structure, often of infinite size (Kleinberg 2000). In our setting, we grapple with dynamic, heterogeneous, real-world networks composed of a finite number of mobile individuals. There is much work to be done on clarifying the relationships between results on navigation found for the canonical networks studied in theory and the navigability of real-world networks.

Limitations

We use the geocoded Twitter data as a proxy for the location and movement of a large number of individuals. Twitter has over 500 million active users who collectively write nearly 9,000 tweets per second, on average.[2] We focus on a subset of Twitter activity that includes hundreds of thousands of individuals who geo-tag their tweets in specific geographical regions (see the Dataset section above). We estimate the performance of our proposed delivery system in which these users participate. While the deliveries are hypothetical, the delivery routes are based on data generated by people as they traverse and tweet from locations in the real world. Statistical analysis shows that the patterns seen in the mobility of the people in the dataset agree with results reported in other domains, where researchers used other proxies (*e.g.*, cellphone tower data, personal GPS loggers, and WiFi connectivity) for people's actual locations. Thus, we believe that the experiments we shall describe below demonstrate the level of performance that we can expect for a crowd delivery service, *assuming* that the users in the dataset who "participated" in the experimental setting would be involved in the actual routing of packages.

[2]http://www.statisticbrain.com/twitter-statistics/

People identified as contributors in our experiments might not agree to participate in a live exercise. Thus, the routing results we describe can be viewed as an upper bound on the performance of a deployed system. At the same time, a live study would provide opportunities for influencing peoples behaviors with incentives, with advantages in routing coming with requested diversions from an intended path, or even with confirmation that people would take their normal expected daily paths. For instance, if Joe is paid \$5, he will be happy to go to work on a Saturday instead of a Friday, thereby increasing the permeability of the system, since there would be otherwise a lack of workers along his commute route on Saturday. As we will see, our experiments explore the influence of *local* perturbations of people's location (*e.g.*, What if Joe went 100 meters south in the next 10 minutes to be able to meet the current package carrier?), but we do not modify intended paths beyond this. Additionally, discrete tweets constitute a very sparse sample of a person's trajectory. A live system might motivate workers to share their location at a higher sampling rate, potentially via an automated heartbeat with known fixed or context-sensitive frequency.

Related Work

We refer to much related work throughout the paper. We discuss in this section the broader context of related research.

Location plays a central role in our lives. With the advent of mobile Internet-enabled devices, fine-grained location data is now captured at a population scale from people who either implicitly or explicitly opt-in to share their data with the public or with service providers in return for enhanced services. This data has enabled multiple recent studies of human mobility that leverage a wide variety of data from GPS loggers, social networks, bank notes circulation records, and call logs (Krumm and Horvitz 2006; Brockmann, Hufnagel, and Geisel 2006; Liao et al. 2007; Song et al. 2010; Cheng et al. 2011; Sadilek, Kautz, and Bigham 2012).

Our work depends on a graph of spatial and temporal connectedness among people. Following the pioneering work of Milgram that gave rise to the notion of "six degrees of separation," researchers have shown that people are heavily connected in a wide range of settings—both offline and online (Milgram 1967; Dodds, Muhamad, and Watts 2003; Adamic and Adar 2005; Leskovec and Horvitz 2008). Remarkably, not only do short paths exist between arbitrary pairs of individuals, but we can also efficiently *find* them (Clauset and Moore 2003; Liben-Nowell et al. 2005; Lattanzi, Panconesi, and Sivakumar 2011).

However, the models are based on a family of random infinite graphs. In the CP domain, we deal with more complex networks: dynamic agents with uncertain location and connectivity, an inhomogeneous finite lattice, and a nonuniform population distribution. While researchers have proposed theoretical formalisms that begin to address some of these real-world challenges (Clauset and Moore 2003; Liben-Nowell et al. 2005), the overall picture remains unclear. We hope that the empirical results presented here will

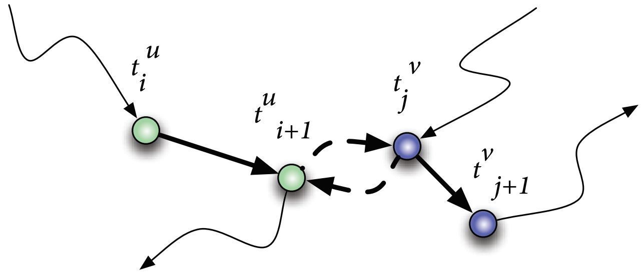


Figure 2: Example of a routing graph G with two people (green u and blue v). Each node represents user location at a given time. Edges induce possible paths a package can follow by connecting consecutive locations of a person (solid edges), and exchange opportunities between people (dashed). All edges are weighted by the time it takes to travel between nodes and by wait time.

inspire additional theoretical efforts to elucidate multiple aspects of crowdphysics challenges.

Prior work on delay tolerant networks explored distributed routing schemes among mobile wireless devices (Fall 2003; Yuan, Cardei, and Wu 2009). CP faces additional challenges, however. Unlike data packets, workers and contiguous physical objects cannot be arbitrarily cached, duplicated, and broadcasted. However, objects too heavy to carry by individuals might be decomposed into lighter components, separately transported to a destination, and then reassembled.

Researchers have begun to explore *multi-agent* crowdsourcing tasks as well (Zhang et al. 2012a; 2012b). For example, the ReGroup system assists in creation of online groups on the basis of individuals' attributes (Amershi, Fogarty, and Weld 2012), and Scribe merges noisy and incomplete speech transcripts produced by a number of workers typing in parallel into a single coherent caption (Lasecki et al. 2012). However, such settings require only synchronization in time, not in physical space.

Related efforts on computation for physical coordination include efforts on ridesharing and on opportunistic planning. The work on ridesharing demonstrates the value of methods that consider divergence in distance and time for participants, and were tested with real-world commuting workloads for global optimization and streaming settings (Kamar and Horvitz 2009). Work on opportunistic planning has centered on the use of predictive models to recommend the injection of waypoints into existing plans opportunistically, considering the divergence from destinations under uncertainty (Horvitz and Krumm 2012).

Methodology and Models

We now provide details about our approach, models, and their application. We shall more crisply define terms that we have been using qualitatively. We start with the dataset described in the previous section, and induce a routing graph from user IDs, locations, and timestamps associated with the tweets. We then use an efficient graph search method to plan globally optimal package routings, which are in turn leveraged to model the expected performance of TwedEx in terms of delivery times, number of routing hops needed, and other statistics of interest. We then contrast this optimal upper bound with the performance of local routing algorithms working under uncertainty and incomplete knowledge about future locations. Finally, we explore mechanisms for a novel form of delivery: delivering to and among people who are on the move.

To quantify the performance of delivery within a geographical area, we divide the area into a uniform grid of 450 by 450 meter cells. *Reachability* (coverage) is defined as the percentage of cell pairs for which a feasible path exists. Delivery time $c_1 \rightsquigarrow c_2$ is equal to the duration of the shortest path of handoffs between the two cells, considering all tweets in c_1 as origins and all tweets in c_2 as targets. (For n cells, there are $n(n-1)$ different ordered pairs.)

Specifically, we explore the following research questions and measures of delivery performance:

1. Coverage and reachability: How large is the geographical area that the system can service, and which combinations of package origin–destination locations are feasible?
2. Delivery time: How long does it take to deliver the package? We measure this period as the time elapsed between pickup and final delivery (timestamps of the first and last tweet on the path, respectively).
3. Optimality: How effective are local routing heuristics working under uncertainty as compared to the optimal solution?
4. Robustness: How do performance metrics change as we strategically remove participants from the network?
5. Locale sensitivity: How do the metrics change across different cities and geographical scales?

We begin with the construction of the routing network.

Routing Graph

From a set of geotagged tweets, we induce a weighted directed graph G—a routing network. The i-th tweet of user u is represented as a node t_i^u in G, and there is a directed edge (t_i^u, t_{i+1}^u) for all pairs of consecutive tweets of user u (see Figure 2). Thus, each user contributes a directed path $(t_1^u, \ldots, t_N^u)$ to G, where N is the total number of tweets written by user u in the data collection period. Every edge (t_i^u, t_{i+1}^u) is weighted by the time elapsed between posting t_i and t_{i+1}: $w_{(t_i^u, t_{i+1}^u)} = \mathrm{time}(t_{i+1}^u) - \mathrm{time}(t_i^u)$.

To model package handoffs, we add edges to the routing graph G that denote the potential exchange of a package between two individuals. Before we do this, however, we need to define a *meeting* between people. In this paper, we say two users meet if they tweet within specified distance and time thresholds. More precisely, users u and v meet if there exist t_i^u and t_j^v, such that $\mathrm{distance}(t_i^u, t_j^v) < \delta$ and $|\mathrm{time}(t_i^u) - \mathrm{time}(t_j^v)| < \tau$. δ and τ are parameters we vary in the experiments below in order to explore the sensitivity of routing performance to increasing the digressions people make from their intended paths or dwell times. We seek to understand how such changes—which might be achieved through selective payments in a real-world system—can lead to routing graphs with target levels of permeability and coverage.

For a given setting of δ and τ, we add to G a set of possible exchange edges

$$\left\{(t_i^u, t_j^v) : \mathrm{distance}(t_i^u, t_j^v) < \delta \wedge \left|\mathrm{time}(t_i^u) - \mathrm{time}(t_j^v)\right| < \tau\right\}.$$

The weight of these exchange edges is equal to the expected wait time or the time it takes to make the required digression, whichever is greater:

$$w_{(t_i^u, t_j^v)} = \max\left(\left|\mathrm{time}(t_i^u) - \mathrm{time}(t_j^v)\right|, \frac{\mathrm{distance}(t_i^u, t_j^v)}{1.4^{m}/_{s}}\right),$$

where we divide the distance covered by the preferred walking speed of $1.4^{m}/_{s}$ to obtain the expected walk time. As a result, all edges in G are weighted by time—measured in seconds—that elapsed between the two incident nodes (tweets). The exchange edge is added only if both users can still arrive at their next location in time.

Global Planning

The induction of G is now complete and we can use the graph to find the optimal delivery route among any pair of locations by simply running any shortest-path search algorithm, for example Dijkstra's algorithm. The routes found this way will be globally optimal, using all the available information, and will be useful to establish an upper bound on the performance of the delivery system. However, in live execution, the system's knowledge will be incomplete: people's future location will be uncertain and a routing service will need to use predictions about peoples future behavior. We explore this scenario in the next section.

As we increase the slack parameters δ and τ for counting meetings between people, G becomes denser. As a result, computing shortest paths becomes computationally more challenging. While the number of nodes in G is independent of δ and τ, it is large to begin with (see Table 1). This makes computation of paths between *all pairs* of nodes difficult. For example, our US dataset spanning only 2 days has over 3 million nodes and 10 million edges even for a modest setting $\delta = 100$ meters and $\tau = 1$ hour.

Because of the addition of the handoff edges, the delivery times computed using G are approximate in some cases. For instance, when we allow a person to wait longer by increasing τ, his arrival at the next location may be delayed. Note that this could impact the estimate of the delivery time only if the person who is about to pick up the package is required to wait, or if a person participates in multiple adjacent handoffs. Allowing this small imprecision enables us to plan all optimal routes efficiently.

The standard Floyd-Warshall shortest-paths algorithm runs in $O(n^3)$ time, where n is the number of nodes in G. For sparse graphs, Johnson's algorithm has a lower, but still daunting, complexity of $O(e^2 \log n + en)$, where e is the number of edges (Cormen et al. 2001). To counteract these difficulties, we apply the PHAST algorithm (Delling et al. 2012). PHAST performs a clever preprocessing step on G after which shortest paths from any node to all other nodes can be computed in $O(n)$.

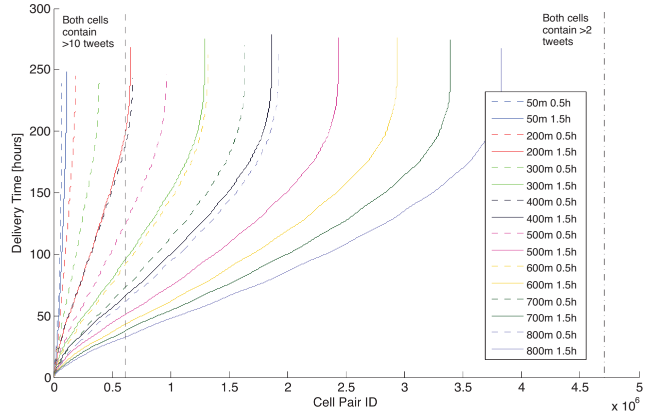


Figure 3: Delivery times among all pairs of cells with sufficient Twitter activity as we vary δ and τ in the Seattle dataset. We see that only modest digressions and wait times are required to cover all cell pairs that contain more than 10 tweets in both cells.

In the following section, we use the shortest paths to quantify an upper bound on a variety of performance measures important for a crowd-powered delivery service.

Local Opportunistic Routing

We seek to quantify the gap between theoretically possible delivery times in our system using global search on retrospective data versus delivery times that can be realistically expected in a live system operating under uncertainty. To do this, we explore the power of employing *local* routing policies that could be executed in real time, relying only on simple statistics about future locations of people computed from historical data. These statistics are then used in heuristics that guide local routing decisions.

We take a large sample of Twitter activity in a 60×60 kilometer bounding box around Seattle over a period of six months. The data is filtered by the same process as described in the Dataset section. We consider the final 35 hours of the data for testing, and all the preceding data for training. The basic idea is to learn a simple model of people's locations, and subsequently leverage the model to make routing decisions using only local information. The model of user activity can be made readily available to each worker.

We first extract the set of possible delivery locations $\mathcal{L}$. These are simply the time-stamped locations of all the tweets in the test set. For each $\ell \in \mathcal{L}$, we construct a ranked list of users based on the distances between them and ℓ. Users who often appear close to ℓ are ranked high, and individuals who spend their time far away are ranked low.

The experimental setup for local routing is the same as the approach we described earlier for global routing, except that the shortest paths are not precomputed. We again divide an area into uniform 450 by 450 meter cells. Then, for each pair of cells (c_1, c_2), we simulate routing of package from c_1 to c_2. Every time a person carrying a package encounters another individual, he needs to make a routing decision. He either keeps the package or hands it off to the other person.

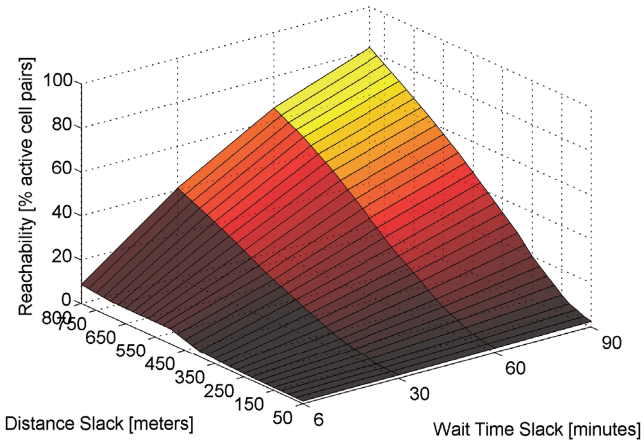


Figure 4: Reachability for all pairs of active cells as we vary the digression slack δ and wait time slack τ in the Seattle dataset, showing phase transition as function of parameters. We say a cell is *active* if it contains at least three tweets.

We use $\delta = 100$ meter and $\tau = 0.5$ hour thresholds to define an encounter. The decision is made based on the ranking of the users M who just met (M includes the current carrier as well). Given that the package's destination is ℓ, the next carrier $u^\star$ is chosen by finding the highest ranked candidate with respect to ℓ in M:

$$u^\star_{\text{closest point}} = \underset{u \in M}{\text{argmax}} \ \text{rank}(u, \ell).$$

Note that this algorithm requires only the knowledge of one's current location, package destination, and a simple statistic extracted from historical data for each person in the vicinity of a routing point.

Experiments and Results

This section reports our experimental results and closely follows the structure of the Methodology and Models section above. For brevity, we will primarily focus on the Seattle dataset, but we will contrast the three locales where interesting differences emerge. Note that Twitter activity is roughly 4.7 times sparser in Seattle than in NYC, across all measures in Table 1. As a result, the delivery performance is significantly higher in NYC. We will explore this difference as well.

Leveraging the grid of 450 by 450 meter cells, TwedEx's coverage of a geographical area is defined as the percentage of cell pairs for which a feasible path exists. Our measure of reachability is strict. If the package ends up only one meter away from the target cell boundary, it is counted as undelivered. A softer metric could be used to take into account the continuous distance, and increasing cost of divergence could be assigned to outcomes with longer distances with respect to the target. However, the strict definition gives us a clean, discrete measure of success.

Additionally, we can remove the constraint that the package has to always be carried by a participant. Designated "geo caches" can be used for safe storage from which the package is retrievable by a code sent to the best candidate for performing the pickup. For clarity, we focus on the discrete definition of reachability and direct handoffs among workers throughout this paper.

Global Planning

Recall that we run PHAST on the routing graph G induced from our data to obtain all pairs of shortest paths among all points in space and time defined by the tweets. We then measure the delivery times, coverage, and number of hops required for each route, and compare the performance of delivery in three locales (NYC, SEA, and US). Delivery time from cell c_1 to cell c_2 is equal to the duration of the shortest path between the two cells, considering all tweets in c_1 as origins and all tweets in c_2 as targets.

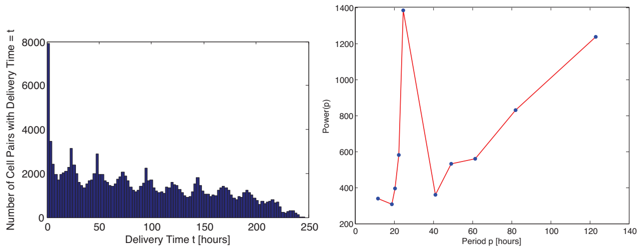
Figure 3 shows delivery times between all pairs of cells as we vary δ and τ in the Seattle dataset. We see that only modest geographic digressions and wait times are required to cover all cell pairs that contain more than 10 tweets in both cells (this threshold is shown with the leftmost straight vertical line). For example, less than a 200 meter digression and less than a 1.5 hour wait time are sufficient to achieve this coverage (solid red curve). $\delta = 400$ meters and $\tau = 0.5$ hours is sufficient as well (dashed black curve).

Reachability—measured by the percentage of cell pairs covered by the system—for all pairs of cells as we vary the digression slack δ and wait time slack τ in the Seattle dataset is displayed in Figure 4. Even large values of δ and τ *individually* do not improve the coverage significantly. However, increasing both jointly results in a *phase transition* at $\delta = 350$ and $\tau = 0.5$ after which the system becomes significantly more permeable and ultimately approaches full coverage with digressions up to 800 meters and wait times bounded by 90 minutes.

Since NYC is considerably denser and has more Twitter activity, it is nearly four times more permeable than Seattle for any setting of δ and τ. We are interested in normalizing the delivery performance in different locales. Therefore, we choose the top-100 cells with the most activity in each respective city and concentrate on reachability and delivery times only for pairs of cells from these sets. We observe that the delivery time distribution among these cells does not significantly differ across locales.

Figure 5a shows a typical distribution over delivery times we find across locales and parameter settings. Looking more closely at the patterns in delivery times, we perform frequency analysis of the distribution using fast Fourier transform (FFT) and find peaks corresponding to exactly 24 hours and 5 days (a work week); see Figure 5. We believe that these results reveal natural periods reflecting the modulation of the lives of people and find it interesting that these natural patterns show up strongly in *joint* human activities.

Delivery time is one measure the efficiency of a delivery system. We are also interested in the number of hops a package makes as it is carried by the workers. A hop is a jump of a package between two consecutive tweets, either with or without changing hands between workers. We find that the distribution is heavily skewed towards small numbers of hops. Specifically, the probability of observing a certain number of hops k decreases exponentially with k.



(a) Delivery time distribution

(b) Power spectrum

Figure 5: (a) Distribution of delivery times among all pairs of origin-destination cells, for $\delta = 100$ and $\tau = 0.5$ in Seattle. A large number of cell pairs are reachable in just a few hours. We find that a disproportionate number of deliveries occurs at integer multiples of 24 hours. (b) Frequency analysis of the distribution using Fourier transform shows peaks corresponding to exactly 24 hours and a work week (5 days).

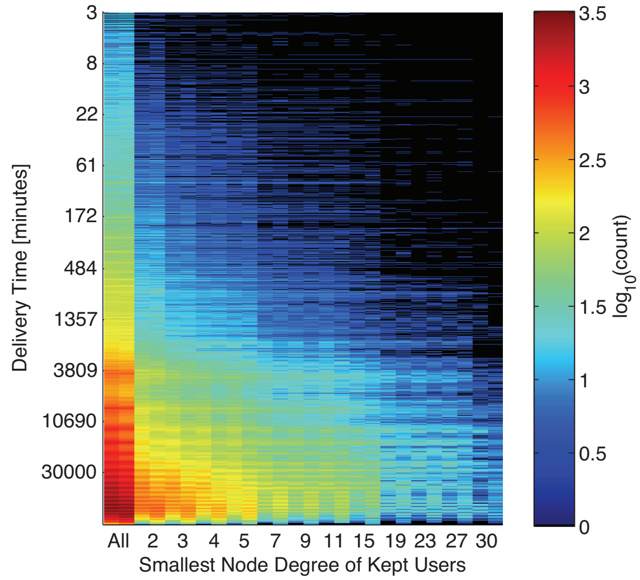


Figure 6: Patterns in delivery times among cell pairs as participants are removed progressively based on their connectedness in the meeting graph.

The most likely interval is $k \in [1, \ldots, 6)$. At the same time, there several instances of a package making more than 400 hops.

We now turn to a study of the robustness of routing performance while removing participants who share certain characteristics. The sensitivity of coverage and performance to removal of workers is shown in Figure 6. A person who met ($\delta = 100$ meters and $\tau = 0.5$ hours) n different people over the observation period has a node degree n. We progressively remove users with degree 1, 2, *etc.*. Interestingly, removing seemingly insignificant workers with degree less than two has a considerable negative influence on TwedEx, both in terms of coverage and expected delivery times. We conclude that the sparse occasional activity of low-rank users is essential for the entire system to be truly permeable.

Figure 7 shows the relationship among geodesic distances between cells scattered across the entire continental United States, and the delivery times between the cells. The results are for global routing with $\delta = 100$ meters and $\tau = 0.5$ hours for handoffs. The color represents the number of occurrences with particular values of distance and delivery time. We see that delivery between nearby cells requires small amounts of time (typically under 5 hours), but even distant locations can be reached in five hours (lower-right). Delivery between many distant cell pairs is possible in an amount of time only slightly higher than it takes to fly directly between the locations.

This holds even for cells that are not near airports and therefore requires a planned route that takes the package to a nearby airport, finds a worker who is about to fly to the target city, and ultimately routes the package to its final non-airport destination. We see that TwedEx takes advantage of people moving around locally and flying around globally to deliver packages quickly. It is not unusual to obtain a 5.5 hour delivery NYC $\rightsquigarrow$ Los Angeles or San Francisco $\rightsquigarrow$ NYC, where the end points of the route are *not* airports.

It is not surprising that local deliveries take less time, but it is interesting to see that some large distances are easier to traverse than others. For example, delivering coast to coast is much faster than between a coast and the center of the US. 2,000 kilometers *from anywhere in the US* seems to be the worst distance to traverse. This number likely arises from the geography of the country, where most people are concentrated along the coasts, the coasts are roughly 4,000 kilometers apart, and each coast is shorter than 2,500 kilometers (north to south).

Local Opportunistic Routing

So far we have focused on routing with global knowledge, assuming there is an oracle that participants consult at each decision point. The oracle has access to the full routing graph and therefore can calculate and provide globally optimal paths for any given pair of source-target locations. This establishes a tight empirical upper bound on the expected performance of a crowd delivery service. However, the performance of such a delivery service running in real time—where future locations are uncertain—remains an open question. Our experiments in this section begin to answer this question in the context of a dynamic heterogeneous system composed of people moving in the real world. Therefore, we now shift our attention to routing under uncertainty, in the absence of deterministic knowledge about future locations.

Figure 8 compares the performance of the local, opportunistic routing algorithm described in the previous section (we call it *closest point* routing) to a global optimum and to a random baseline. The global optimum is determined by finding shortest paths between all pairs of cells by executing PHAST algorithm on a fully observed routing graph G. The random baseline simply chooses the next carrier $u^\star$ uniformly at random from the set of available workers M at

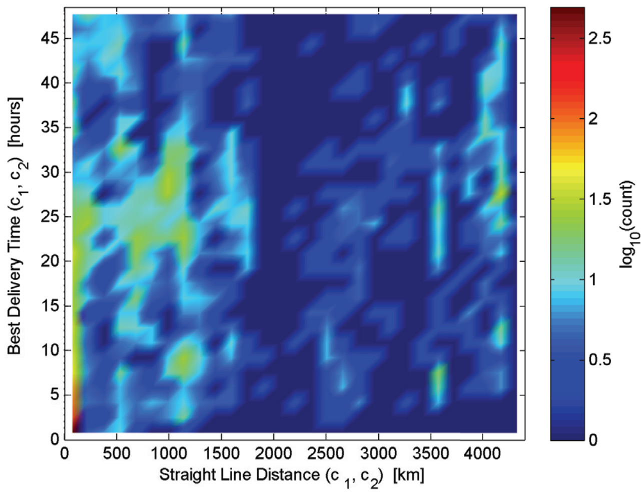


Figure 7: Relationship between the geodesic distances among cells scattered across the entire continental US, and the delivery times between the cells.

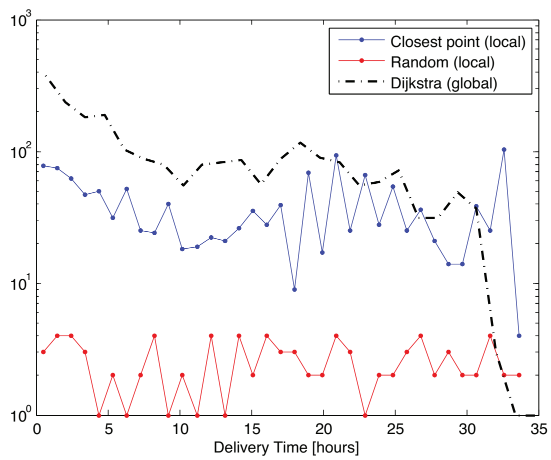


Figure 8: Comparison of coverage and delivery times for two local routing algorithms to a globally optimal performance (PHAST implementation of Dijkstra) for parameter settings $\delta = 100$ meters and $\tau = 0.5$ hours.

each routing point:

$$u^{\star}_{\text{random}} \sim \mathcal{U}(1, |M|).$$

We see that the closest-point heuristic results in a distribution of delivery times that is slightly skewed towards longer times, but it closely follows the optimal distribution. Closest-point routing achieves 58% of the coverage attained by globally optimal routing, whereas the random strategy achieves only 4% of the coverage. To our knowledge, this is the first empirical result on local routing under uncertainty in dynamic networks composed of mobile individuals. We find that, even using very simple heuristics, such as aiming for the person who is most likely to appear near the package's destination given historical data, is effective.

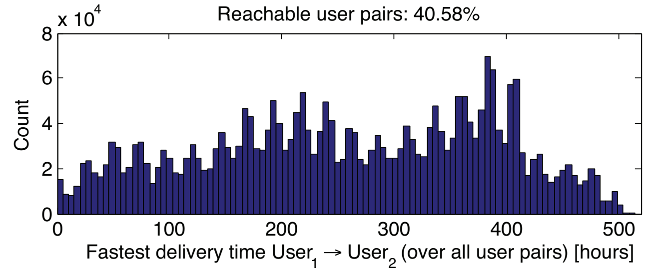


Figure 9: Distribution over delivery times when setting *people* as package "destinations". This graph is for our Seattle dataset. We see that for roughly half of the people-pairs, a delivery path that takes less than 10 days exists and can be found. As in our previous results for delivery times to fixed locations, we identify a strong periodicity in the system.

Delivery to a Moving Target

Finally, the distributed nature of crowd physics enables services that are currently unavailable, such as delivery to a specific person who is on the move. Instead of routing to a fixed street addres—the singular goal of todays delivery services—we can leverage the crowd to deliver to a *moving* target. While expensive courier services will deliver to a specific place (not necessarily one's home address) at a specific time, we are interested whether the crowd can provide more flexible delivery to people in motion *at scale*. Imagine that you wish to send a gift to a friend. You write only the friends unique identifier on the package and drop it into the crowd substrate. The identifier can be a telephone number, email address, or Twitter handle. Workers then route the package among themselves until one of them meets the target person. Thus, the only constraint in this system is that the package has to be delivered to the recipient as fast as possible. Aside from that, the delivery can occur anywhere, anytime. This is somewhat analogous to sending a text message, which quickly reaches the recipient's mobile phone regardless of location. However, since packages are physical objects and we do not know people's location ahead of time, things are more complicated.

For all users u, we take u's first tweet that appears in the dataset (t^u_1). For all other users v, we then find the fastest path $u \rightsquigarrow v$ through the routing network G as we have been doing all along, except v is now a moving target. This is done for all pairs of users

$$\{(u, v) : u, v \in U \wedge u \neq v\},$$

where U is the set of all users in the dataset. Delivery is considered successful if there exists a path $u \rightsquigarrow v$ and the delivery time is measured from t^u_1 to the first encounter of u and v, where the package would have been delivered. If no such path exists, we say u cannot reach v. Since u is defined only by their first tweet t^u_1, this method also shows the reachability of people from locations, namely all locations associated with the set of tweets $\{t^u_1 : u \in U\}$.

The histogram of delivery times that arise in this setup with global routing is shown in Figure 9 for the Seattle dataset with $\delta = 100$ meters and $\tau = 0.5$ hours (handoffs are possible only if workers are no more than 100 meters and 30 minutes apart). Even though the meeting matrix is sparse, we can reach 41% of people-pairs. Again, we observe strong periodicity in the delivery times with large 24 hour and 7 day peaks in the FFT power spectrum.

Conclusions and Future Work

We introduce and explore a rich class of *crowdphysics* problems, which contains tasks that require coordination of people in space and time, as distinct from existing crowdsourcing research. We focus on a distributed package delivery task as a representative instance of crowdphysics. We propose and evaluate two approaches to route planning in this domain, using geotagged tweets as a proxy for people's locations. We consider ideal global coordination using retrospective data, as well as local opportunistic routing under uncertainty. Both approaches are expressed in a unified way via a reduction to a graph search problem that can be solved efficiently.

We find that delivery can have remarkable speed and coverage. For example, with digressions δ smaller than 800 meters and dwell times τ shorter than 90 minutes, the delivery service covers 83% and 100% of source-target location pairs within the greater Seattle and New York City metropolitan areas, respectively. Even tight bounds on deviations result in good coverage. For example, 100 meters and 30 minutes is enough to cover 18% of source-origin location pairs in Seattle and over 50% in NYC. We further show that even poorly connected people in the network play, in aggregate, a major role in the permeability of the system.

Similar results on reachability and performance hold even when routing without global knowledge, relying only on local greedy decisions. To our knowledge this is the first empirical evidence that dynamic networks composed of people in motion are highly navigable. Finally, we examine services that are unavailable today, such as delivery to a specific person (unconstrained by a street address) in a scalable way. We show that a large fraction of package transmission among all people-pairs can be completed within 21 days.

Future directions of research include exploring an array of local decision policies, including routing based on the composition of mixtures of factors weighted by parameters learned from data. For example, an occasional biased randomization could lead to better performance of local navigation. We seek to better understand how a computed global optimum might be surpassed with clever management of the crowd with methods for learning predictive models and decision making, as has been shown in other domains (Kamar, Hacker, and Horvitz 2012).

Another direction is to aggregate probabilities on transitions between nodes in our graph. This invites the use of established theoretical results with random walks on graphs for evaluating probabilistic expectations on delivery performance, in terms of hitting times and stationary distributions.

Trust, motivation, incentives, and payments are important considerations in crowdphysics problems. For example, a sender may wish to constrain a delivery service to only route between pairs of people who are within a certain distance in the friendship graph, or assert trust requirements among all people composing the chain of participants that touch a package. Alternately, the system may learn on its own about the trust relationships required to achieve levels of reliability in safe and efficient transport. Such preferences will likely need to be traded against speed of delivery; it will be interesting to see which aspects of the structure of the underlying routing network lead to increased robustness and permeability, per consideration of critical components of the graph, key connectors, and essential backbone of participants.

Beyond relying on assessed social connections, signals in location data can be used to quantify the expected level of trust among individuals. For example, two people who appear at several *different* venues together are often related in some way (Crandall et al. 2010; Sadilek, Kautz, and Bigham 2012). Regarding incentives and payments, promising research includes the study of solutions that allow for a trading off of quality of solutions and the total cost required for issuing sets of payments, where targeted payments reimburse participants for real-time or preplanned divergences in space and time.

In other work, we are interested in the value of taking a computational perspective on challenges in epidemiology, seeking to understand how results on reachability and performance in navigating graphs might be used to extend traditional epidemiological approaches to modeling disease transmission. As we mentioned earlier, computational procedures for designing ideal disruption of the spread of illness may yield new approaches in public health (Sadilek and Kautz 2013).

We are interested in further exploring multiple scenarios for engaging people to perform coordinated sensing and action in the world. The applications span a broad spectrum of tasks beyond package delivery, including search and rescue, epidemiology, sensing, creation of structures, and even execution of large-scale social and political activities. There are numerous opportunities to leverage human computation, optimization, and machine learning to develop live crowdphysics solutions that bring prototypes like TwedEx to life. Constructing and fielding these services will undoubtedly provide new insights.

Acknowledgments

We thank Daniel Delling for his help with running PHAST. We further thank Walter Lasecki, Alexander Holroyd, and Yuval Peres for interesting discussions and pointers.

References

Adamic, L., and Adar, E. 2005. How to search a social network. *Social Networks* 27(3):187–203.

Amershi, S.; Fogarty, J.; and Weld, D. 2012. Regroup: Interactive machine learning for on-demand group creation in social networks. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, 21–30. ACM.

Barabási, A., and Albert, R. 1999. Emergence of scaling in random networks. *science* 286(5439):509–512.

Bernstein, M.; Little, G.; Miller, R.; Hartmann, B.; Ackerman, M.; Karger, D.; Crowell, D.; and Panovich, K. 2010. Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd annual*

ACM symposium on User interface software and technology, 313–322. ACM.

Bigham, J.; Jayant, C.; Ji, H.; Little, G.; Miller, A.; Miller, R.; Miller, R.; Tatarowicz, A.; White, B.; White, S.; et al. 2010. Vizwiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, 333–342. ACM.

Brockmann, D.; Hufnagel, L.; and Geisel, T. 2006. The scaling laws of human travel. *Nature* 439(7075):462–465.

Cheng, Z.; Caverlee, J.; Lee, K.; and Sui, D. 2011. Exploring millions of footprints in location sharing services. *AAAI ICWSM*.

Clauset, A., and Moore, C. 2003. How do networks become navigable? *arXiv preprint cond-mat/0309415*.

Cormen, T.; Leiserson, C.; Rivest, R.; and Stein, C. 2001. *Introduction to algorithms*. MIT press.

Crandall, D.; Backstrom, L.; Cosley, D.; Suri, S.; Huttenlocher, D.; and Kleinberg, J. 2010. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences* 107(52):22436.

Delling, D.; Goldberg, A.; Nowatzyk, A.; and Werneck, R. 2012. Phast: Hardware-accelerated shortest path trees. *Journal of Parallel and Distributed Computing*.

Dodds, P.; Muhamad, R.; and Watts, D. 2003. An experimental study of search in global social networks. *Science* 301(5634):827–829.

Eubank, S.; Guclu, H.; Anil Kumar, V.; Marathe, M.; Srinivasan, A.; Toroczkai, Z.; and Wang, N. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429(6988):180–184.

Fall, K. 2003. A delay-tolerant network architecture for challenged internets. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, 27–34. ACM.

Faloutsos, M.; Faloutsos, P.; and Faloutsos, C. 1999. On power-law relationships of the internet topology. In *ACM SIGCOMM Computer Communication Review*, volume 29, 251–262. ACM.

González, M.; Hidalgo, C.; and Barabási, A. 2008. Understanding individual human mobility patterns. *Nature* 453(7196):779–782.

Horvitz, E., and Krumm, J. 2012. Some help on the way: Opportunistic routing under uncertainty. In *UbiComp 2012*.

Kamar, E., and Horvitz, E. 2009. Collaboration and shared plans in the open world: Studies of ridesharing. In *IJCAI*.

Kamar, E.; Hacker, S.; and Horvitz, E. 2012. Combining human and machine intelligence in large-scale crowdsourcing. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, 467–474. International Foundation for Autonomous Agents and Multiagent Systems.

Kleinberg, J. 2000. Navigation in a small world. *Nature* 406(6798):845–845.

Krumm, J., and Horvitz, E. 2006. Predestination: Inferring destinations from partial trajectories. *UbiComp 2006: Ubiquitous Computing* 243–260.

Lasecki, W. S.; Miller, C. D.; Sadilek, A.; Abumoussa, A.; Borrello, D.; Kushalnagar, R.; and Bigham, J. P. 2012. Real-time captioning by groups of non-experts. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12.

Lattanzi, S.; Panconesi, A.; and Sivakumar, D. 2011. Milgram-routing in social networks. In *Proceedings of the 20th international conference on World wide web*, 725–734. ACM.

Leskovec, J., and Horvitz, E. 2008. Planetary-scale views on a large instant-messaging network. In *Proceeding of the 17th international conference on World Wide Web*, 915–924. ACM.

Leskovec, J.; Kleinberg, J.; and Faloutsos, C. 2005. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, 177–187. ACM.

Liao, L.; Patterson, D.; Fox, D.; and Kautz, H. 2007. Learning and inferring transportation routines. *Artificial Intelligence* 171(5-6):311–331.

Liben-Nowell, D.; Novak, J.; Kumar, R.; Raghavan, P.; and Tomkins, A. 2005. Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the United States of America* 102(33):11623.

Little, G., and Miller, R. 2006. Translating keyword commands into executable code. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, 135–144. ACM.

Mandelbrot, B. 1982. *The fractal geometry of nature*. Times Books.

Milgram, S. 1967. The small world problem. *Psychology today* 2(1):60–67.

Oliveira, J., and Barabási, A. 2005. Human dynamics: Darwin and einstein correspondence patterns. *Nature* 437(7063):1251–1251.

Sadilek, A., and Kautz, H. 2013. Modeling the impact of lifestyle on health at scale. In *Sixth ACM International Conference on Web Search and Data Mining*.

Sadilek, A.; Kautz, H.; and Bigham, J. P. 2012. Finding your friends and following them to where you are. In *Fifth ACM International Conference on Web Search and Data Mining*. (Best Paper Award).

Sadilek, A.; Kautz, H.; and Silenzio, V. 2012. Predicting disease transmission from geo-tagged micro-blog data. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Shahaf, D., and Horvitz, E. 2010. Generalized task markets for human and machine computation. AAAI.

Song, C.; Qu, Z.; Blumm, N.; and Barabási, A. 2010. Limits of predictability in human mobility. *Science* 327(5968):1018.

Von Ahn, L., and Dabbish, L. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 319–326. ACM.

Wu, F.; Huberman, B.; Adamic, L.; and Tyler, J. 2004. Information flow in social groups. *Physica A: Statistical Mechanics and its Applications* 337(1):327–335.

Yuan, Q.; Cardei, I.; and Wu, J. 2009. Predict and relay: an efficient routing in disruption-tolerant networks. In *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*, 95–104. ACM.

Zhang, H.; Horvitz, E.; Chen, Y.; and Parkes, D. 2012a. Task routing for prediction tasks. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, 889–896. International Foundation for Autonomous Agents and Multiagent Systems.

Zhang, H.; Law, E.; Miller, R.; Gajos, K.; Parkes, D.; and Horvitz, E. 2012b. Human computation tasks with global constraints. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, 217–226. ACM.