The limits of social mobilization

Alex Rutherford [*], Manuel Cebrian [†] [‡], Sohan Dsouza [*] , Esteban Moro [§], Alex Pentland [¶], and Iyad Rahwan [1] [*] [‖]

[*]Computing & Information Science, Masdar Institute of Science and Technology, Abu Dhabi, UAE,[†]Department of Computer Science & Engineering, University of California at San Diego, La Jolla, California, USA,[‡]NICTA, Melbourne, Victoria, Australia,[§]Department of Mathematics, Universidad Carlos III de Madrid, Madrid, Spain,[¶]Media Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, and [‖]School of Informatics, University of Edinburgh, Edinburgh, UK

Submitted to Proceedings of the National Academy of Sciences of the United States of America

[1]To whom correspondence may be addressed. Email: irahwan@acm.org

The Internet and social media has enabled the mobilization of large crowds to achieve time-critical feats, ranging from mapping crises in real-time, to organizing mass rallies, to conducting search-and-rescue operations over large geographies. Despite significant success, selection bias may lead to inflated expectations of the efficacy of social mobilization for these tasks. What are the limits of social mobilization, and how reliable is it to operate at these limits? We build on recent results on the spatio-temporal structure of social and information networks, to elucidate the constraints they pose on social mobilization. We use the DARPA Network Challenge as our working scenario, in which social media was used to locate 10 balloons across the United States. We conduct high-resolution simulations for referral-based crowdsourcing and obtain a statistical characterization of the population recruited, geography covered, and time to completion. Our results demonstrate that the outcome is plausible without the presence of mass media, but lies at the limit of what time-critical social mobilization is capable of. Success relies critically on highly connected individuals willing to mobilize people in distant locations, overcoming the local trapping of diffusion in highly dense areas. Yet, even under these highly favorable conditions, the risk of unsuccessful search remains significant. These findings have implications on the design of better incentive schemes for social mobilization. They also call for caution in estimating the reliability of this capability.

social mobilization | networks | search

The Internet and online social media are now credited with the unprecedented ability to coordinate the mobilization of large masses of people to achieve incredible feats that require coverage of large geographical and informational landscapes in very limited time. Social media has been used to mobilize volunteers to map natural disasters in real-time [1], and to conduct large-scale search-and-rescue missions [2]. Online social networks have also been an important tool in the coordination of mass political rallies [3, 4].

Endeavors like the DARPA Network Challenge [5] aimed to test the power of the Internet and social media in time-critical social mobilization to its absolute limits. The Network Challenge required competing teams to locate and submit the coordinates of 10 tethered weather balloons dispersed at random locations all over the continental United States. The winning team, based at MIT, won the challenge by locating all balloons in less than 9 hours. The MIT team used an incentive scheme to kick start an information and recruitment cascade that resulted in 4,400 sign-ups to the team's Web site within 48 hours. Analysis of the diffusion revealed that the recursive incentive scheme may have played an important role in maximizing the speed and branching of the diffusion to limits above what is normally observed in viral propagation schemes [6–8].

More recently, the State Department's Tag Challenge required competing teams to locate and photograph 5 target "thieves" (actors) in 5 different cities in the US and Europe, based only on a mug shot released at 8:00am local time [9]. The targets were only visible for 12 hours, and followed normal itineraries around the cities of Stockholm, London, Bratislava, New York City and Washington D.C. Our winning team located 3 of the five suspects using social media, without any of the team members being based in any of the target cities [10], demonstrating yet another example of time-critical social mobilization in tasks that require covering large geographies.

Despite these numerous successes, we still have limited understanding of the limits of technology-mediated mobilization. If we are to rely on social media to react to time-critical emergencies, it is important to understand the conditions under which they can be successful, and the risks of failure associated with them. A particular case, of highly practical importance, is to understand the extent to which we can expect to cover a certain geographical area in a given amount of time. For this, we must understand the complete statistical characterization of the population recruited, geographical area covered, and completion time it takes for social mobilization to succeed in a particular task, as well as to quantify the likelihood of failure.

This lack of understanding is especially prone to selection bias over few successful social mobilization strategies and may lead to inflated expectations of the reliability and efficacy of of these techniques [11, 12]. Yet it is beyond experimental capabilities to perform randomized experimentation with large crowdsourcing challenges (with notable exceptions emerging recently [13, 14]).

Modeling efforts in the wake of the H1N1 and other global pandemics have also provided a valuable insight into time sensitive human dynamics on a large scale, via spatial simulation [15] or network based diffusion [16, 17]. In common with these efforts we model the interaction and connection of large numbers of agents, however we consider the propagation of a *message* which may be transmitted without direct physical proximity and generally shorter 'incubation' times leading to faster spreading. Thus the mechanism of 'infection' (recruitment) is independent of human mobility patterns [18–21], which in our case contribute only to the area searched.

In this work, we build on recent results on social network structure, information diffusion and urban economics, to elucidate the constraints that they pose on social mobilization. In particular, we conduct high-resolution simulations of the DARPA Network Challenge. We obtain statistical characterizations of the population recruited, geography covered, and time to locate the 10 balloons, together with their dependencies on the instrumental variables.

Our results demonstrate, surprisingly, that the DARPA Network Challenge outcome is plausible, and thus it is not simply a fluke that can only be explained by the role of mass media. Having said that, the challenge lies at the limit of what time-critical social mobilization is capable of. Mobilization requires highly connected, highly active individuals to be motivated to propagate the message to a large number of friends, and to mobilize people in distant locations, overcoming

Reserved for Publication Footnotes

the local trapping of diffusion in highly dense areas. Moreover, even under highly favorable conditions, the risk of mobilization failure remains significant. These findings have implications on the design of better incentive schemes for social mobilization. They also call for caution in estimating the reliability of this capability.

Simulation Model

In seeking to understand social mobilization we must consider the many different dynamics that underpin such a process; in particular the branching dynamics of recruitment, the temporal dynamics of message propagation, the geographical spread of social networks, and the scales and aspects of human mobility. A full accountability of each process will render the simulation and its understanding impossible and thus we concentrate on the main ingredients that explain the observed behavior in each of the processes.

Recruitment. Examination of the dynamics of the branching recruitment process in empirical data from the challenge [6] suggests several key features. After a large initial round of recruitment from the seed node, the reproductive number is well below the tipping point (see SI Appendix). Thus a large number of sub-trees are first created spreading from the root node which then steadily die out. In order to describe the typical branching recruitment process, we fit to the observed branching data assuming an atypical burst of recruitment when the search commences. We find a power law distribution with a mean $\langle R_o \rangle = 0.89$. (See Materials and Methods for details of fit).

Response Time. The importance of the heterogeneity in response times in viral recruitment processes has been demonstrated [22]. In a study of a viral email campaign, the time taken to forward a message was found to be log-normally distributed as opposed to the commonly used Gaussian assumption, with a mean of 1.5 days with a standard deviation of 5.5 days. This large heterogeneity has a deep impact on the propagation of information: cascade dynamics may be halted by the few individuals with very long response times and thus recruitment events may continue up to the order of years after the seed node starts the cascade. The waiting time distribution in a time-critical campaign such as the DARPA Network Challenge might differ fundamentally since it must necessarily end by a fixed deadline regardless of whether the campaign is successful or not. However we use the distribution of [22] as a reasonable approximation. We do not apply a cut-off at large times, although the tail of the distribution may be effectively truncated since a search may terminate if all balloons are found before recruits with waiting times drawn from the tail of the distribution are able to act. The role of burstiness in diffusion in temporal networks was investigated systematically in [23] by characterising tie strength due to both topological and temporal characteristics. Paradoxically, burstiness was found to promote efficient diffusion at small scales but to hinder it on large scales.

Geography of Ties. Several studies have been made of geographical scaling laws for friendship [24–26]. Liben-Nowell *et al* analysed a blogging network and the relationship between friendship and distance. They concluded that friendship correlates more strongly with a person's *rank*, a measure of the number of closer people, than simply with the distance between people [27].

$$P_{ij} \propto \frac{1}{\sum_{k:r_{ik}<r_{ij}} p_k} \qquad \textbf{[1]}$$

Where P_{ij} is the probability of friendship between agents in two distinct grid cells i and j, p_k is the population in cell k and r_{ik} is the distance from cell i to cell k. The quantity on the right hand side is the *rank* of an agent in i with respect to agent in j, it is a measure of the number of people located between i and j. Thus the spatial distribution of a person's friends is now strongly dependent on the local population density; with the effect that two people separated by a given large distance are more likely to be friends in a rural region than a dense, urban environment. It was also noted that friends could be classified into two distinct types; firstly rank-based friends chosen due to geographical proximity (i.e. sharing a common workplace) in accordance with rank scaling. They also observed a 'background' probability that an agent may be friends with any other randomly chosen agent from across the country, in this case friendships are independent of geography. Further, these two types of friends were found to exist in a ratio of 2.5 distance independent friendships to 5.5 rank-based friendships among the user average of 8 friends. In our simulation, we apply this model of friendship to high-resolution population density data derived from census data [28] (see Materials and Methods).

Passive Recruitment. In addition to the branching, temporal and friendship mechanisms above, we investigate the role of two other mechanisms: passive recruitment, and mobility. We describe these in turn below. The data collected during the DARPA network challenge recorded each person who officially registered with the MIT team allowing them to recruit others and to report findings. However this is only a subset of all the people who became aware of the search; the record of sign-ups gives a measure of the number of new recruits which each individual successfully invites, but not the larger hidden network of individuals who search but do not sign-up or recruit others. We refer to this process as passive recruitment, quantified by the number of passive recruits per individual n_{pass}. There was a considerable number of single nodes reporting findings directly; separate from any recruitment tree (5 of the 10 balloons were reported in this way). The reported traffic to the MIT team's website of 100,000 individuals given only 4,400 signups is further evidence of an unreported, hidden network. This suggests that in addition to the observable chain of individuals which actively recruit others after being recruited themselves, there is a supplementary process whereby individuals become aware of the search, and the associated incentives and will report any balloons in their vicinity, yet are not sufficiently motivated to recruit others. This may be due to low affinity with the potential parent node from which they became aware of the search [29–31]. The effect of mass media and possibly word of mouth would also manifest itself in a similar way. By definition the participation of these individuals is difficult to measure unless they report a balloon, but given the large number of submissions attributed to single nodes which were not part of a recruitment chain, we expect that a sizable number of passive recruits were also participating in the search. This process gives rise to an interesting multiplicative factor, separate from the exponential growth of recruits due to branching.

While the number of passive recruits a person is able to mobilize is intrinsically hard to quantify, a good measure of this number is the number of friends of a user of a typical social networking service such as Facebook. The average degree of the entire global network is around 200 with a large range, but it is observed to be up to 400 amongst the most active users [32]. It is these users which have been observed to drive such viral recruitment processes [22]. As discussed in the SI Appendix, we study this parameter for a wide range of passive recruits: however, since we aim to test the behavior of successful social mobilizations we set it to that upper limit of 400 friends. Note also that maintaining a large social network requires a high level of activity [32]. Thus, by selecting this level of passive recruitment we ensure that those users are also the most "temporally active" population.

Mobility. Census data provides a record of where individuals live, but limiting an individual's effective search area to their home ignores their ability to search their vicinity due to their mobility. Due to the high resolution of the simulations (1km^2), it is fair to expect that recruits will instantaneously find a balloon in their own cell. However it is likely that agents will be mobile during the course of a search allowing them to locate balloons in nearby cells. We quantify this with a radius of gyration (r_{mob}). The realistic modeling of individual mobility patterns on short timescales ($\approx 10^1$hours) is non-trivial. These patterns have a proven seasonal nature due to commuting patterns, Circadian rhythms [33] and friendship [34], but an exact individual agent-scale model would require a complex probabilistic treatment and to account for differences in mobility between rural and urban areas [35] putting it beyond the scope of this model. Therefore we define a fixed *mobility radius* allowing agents to locate balloons within a neighborhood of size r_{mob}. The radius of gyration has been investigated extensively using mobile phone data, although typically these studies have focused on the statistical properties of mobility over the course of weeks and months. However a recent study found that on timescales appropriate for time-critical social mobilization (i.e. up to 12 hours) radii of gyration reached 1-2km [36], with a large range. Since a large spread in radii around the mean is expected on this timescale, we also investigate radii in the range 0-5km in our simulations. This parameter also assimilates other mechanisms such as recruited agents becoming aware of a balloon via face-to-face, word of mouth communication.

In light of recent results which find a variability in mobility radius with respect to rank [37] and population density [38], we investigate a variable mobility radius in inverse proportion to local population density (see SI Appendix). Despite the number of passive recruits being unknown, it is likely that the number of passive recruits would follow a distribution since the number of *active* recruits in the branching recruitment process demonstrated a large range. Therefore we also investigate the affect of a distribution of passive recruits. Our findings are insensitive to the introduction of both a variable mobility radius and a distribution of passive recruits, see SI Appendix.

Results

DARPA Balloon Challenge Feasibility. We conducted 500 searches for the 10 DARPA network challenge balloon locations using parameters of $r_{mob} = 2km$ and $n_{pass} = 400$. We find a success rate of 89%. A large variation is seen in completion times (Fig (1) main plot), however the median completion time amongst successful searches was 2.3 days demonstrating a remarkable agreement with the observed time of 48 hours between beginning recruitment and completion [6]. The combined effect of the heavy tailed distribution for branching factor and large heterogeneity in response time gives rise to a large spread in the time for the pure branching process to terminate [22]. Successful searches terminate upon completion which naturally leads to a completion time distribution which is truncated with respect to the underlying distribution of termination times of the pure branching process. It is against this 'natural' range of termination times of the branching (inset of Figure (1)) that the truncated distribution of completion times for *successful* search must be compared. The full range of parameters is investigated in the SI Appendix. We find that minimum values for mobility radius and passive recruits of 2km and 200 respectively are required for a reasonable level of success.

General Balloon Locations. We investigate the hypothesis that the specific balloon locations chosen in the DARPA balloon challenge contributed positively to the speed with which the balloons were found. We randomly choose cells uniformly sampling a large range of population. Further, we simulate the search for a single balloon in each simulation in order to clearly isolate the effect of balloon location on the number of recruits needed to locate it. Figure (2) shows a plot of the number of recruits needed to locate a balloon as a function of the population density of the balloon cell. While the plot contains some noise, there is a clear trend both that balloons in sparsely populated areas require significantly more people to find, and are less likely to be found at all, compared to those in well populated areas. This is due to a combination of effects, a cell containing fewer potential recruits will more likely be searched at a later time. However this is exacerbated by the fact that the population is far from homogeneously distributed, demonstrating strong spatial auto-correlation (see SI Appendix). Rather a sparsely populated cell is likely to be surrounded by other sparsely populated cells, thus there are considerably less opportunities for recruitment into that cell from its neighbors. Conversely, well populated cells in urban areas experience the opposite effect. We have highlighted the extent to which a balloon becomes more easily found as it is moved to a location with higher density. In this context it can be seen that a few of the chosen balloon locations were in challenging locations, but that overall success is expected.

Searchabilty, Blendability, and Findability. In order to draw more general conclusions about the probability of searching a location, we move away from the specific balloon locations. We can now measure the ease with which every single cell may be searched over the course of many different search realisations. With this in mind, we map the *searchability* (s) of each cell i as

$$\mathrm{s}_i = \frac{n_i^{\mathrm{searched}}}{N}, \qquad [2]$$

where n_i^{searched} is the number of instances in which someone is recruited in cell i out of N searches ($N = 10,000$ for the following results). We see (Fig. (3)) that cells located in dense metropolitan areas are easily searchable as there are many more potential searchers to recruit in those cells, whereas the opposite is true for sparsely populated areas. Figure (3, black points) demonstrates this saturating trend above cell population $\approx 10^4\mathrm{km}^{-2}$. This is far from a linear mapping, as some places are highly searchable despite having only intermediate population. Adding more people to a cell located in a small town increases the searchability a great deal, however the payoff for adding more people to a cell in a large city is negligible.

Intuitively, we could also expect an added difficulty in locating a target in a region of high population density such as Manhattan, despite its density-driven high searchability. We model this difficulty to successfully locate a target at a given place by the *blendability* b_i of a cell i. There are (at least) two distinct sources for this difficulty. Firstly as a characteristic of the city itself: The increased density leads to increased complexity of the physical urban environment [45–47] providing more possibilities for a target to be concealed (e.g. an adobe house in Santa Fe, New Mexico vs. a skyscraper in Manhattan). The other contributor to the degree of blendability of a location comes from the individual perspective: sensory overload in busy places, leading to inattentional blindness [39–41], diminished feelings of individual responsibility to report sightings in large crowds [42, 43], and/or reduced cognitive processing ability due to stress [44]. In all the above cases we can safely infer that the larger the population of a cell p_i, the larger its blendability b_i. We assume that $\mathrm{b}_i \sim \mathrm{p}_i^{\beta}$ similar to how other urban indicators scale with population [48, 49]; e.g. wages and crime with $\beta = 1.25$. We also consider walking speed with $\beta = 1$ in SI Appendix.

Thus, if we define the blendability per person we obtain

$$\hat{\mathrm{b}}_i = \frac{\mathrm{p}_i^{\beta}}{\mathrm{p}_i} = \mathrm{p}_i^{\beta-1}, \qquad [3]$$

and we rescale $\{\hat{\mathrm{b}}_i\}$ to lie in the range $[0, 1]$.

The tension between the searchability and blendability of places is modeled by the *findability* per cell i

$$\mathrm{f}_i = \frac{\mathrm{s}_i}{\hat{\mathrm{b}}_i} \qquad [4]$$

Again we scale $\{\mathrm{f}_i\}$ to be in the range $[0, 1]$, and plot equation [4] as the red points in Figure (3) for $\beta = 1.25$ (we repeat the analysis using $\beta = 1$ in the SI). We isolate a regime of high findability defined by a value greater than 0.8, which corresponds to the grey shaded region, with a population density in the range of [1,100 - 13,500km^{-2}]. We emphasise that the exact findability threshold is not important as the blendability is only defined up to a constant. Comparing Midtown Manhattan (population density 36,627km^{-2}) with nearby Asbury Park, New Jersey (population density 4,975km^{-2}) we see (Figure 3) that, counterintuitively, it may be easier to hide in the former than in the latter. The origin of this result is that if $\beta \geq 1$, then for a large p_i the rate at which searchability increases with population is insufficient to overcome the rate at which blendability increases with population, and thus the findability is maximized in places of intermediate density (this happens when $\hat{\mathrm{b}}$ is an increasing function of p_i; a detailed derivation of the condition for β to display this behavior can be found in SI Appendix).

Finally in Figure (4) we visualise the variation of the searchability, blendability and findability in the vicinity of Manhattan and Asbury Park (see SI Appendix for a full map of the continental US). Manhattan has extremely high population density (strong red shading in upper circle), leading to high searchability. However this is again counteracted by a very high blendability resulting in a relatively lower findability than intuitively expected (medium shading in findability map). In contrast the intermediate population density in Asbury Park leads to a fairly high searchability (medium shading in lower circle). But since the blendability is very low (blue shading of blendability map) the findability is very high. In general it can be seen that areas of intermediate population have high findabilities.

Discussion

Our goal is to understand the practical limits of time-critical social mobilization, and to do so in light of contemporary wisdom about the factors that may affect it: the structure and geographical distribution of social ties, the branching and temporal dynamics of information diffusion via social media, and urban economics. Where possible, we used parameters measured from large-scale empirical results, in order to create a realistic, high-resolution simulation of a mobilization scenario akin to the DARPA Network Challenge.

The popular reaction to the DARPA Network Challenge was that it would be impossible without mass media. Our main finding is that success is actually expected with only social media and under realistic parameters. Assuming an initial burst of motivated individuals, success takes place despite the branching factor being lower than the critical point.

Having said that, we find two sobering and instructive qualifiers. Firstly, despite the average completion time coinciding with the experience in the actual challenge, the long tail distribution of completion time suggests that the risk of failing to locate the targets within a short time-frame is also significant. The second important qualifier is that the challenge lies at the limits of what social mobilization is able to achieve. Success relies on all parameters being at their practical limits: you need highly connected individuals to be motivated to propagate the message to a large number of friends, and to mobilize people in distant locations, overcoming the local trapping of diffusion in highly dense areas.

Our results have implications on the use of social mobilization to achieve time-critical tasks, like mapping crises in real-time, or conducting search-and-rescue operations over large geographies. Novel mobilization mechanisms need to focus on incentivizing those elements of the network that are most conducive to successful mobilization: highly-connected people, with distant friends, and rapid reaction time. These characteristics can be exploited in a new measure of *influence*. One can envisage variants of the winning team's recursive incentive strategy that provide network centrality, distance and/or time-sensitive rewards to recruit such influentials.

We studied tension between the benefits and difficulties of searching for physical objects in highly populated areas by defining measures of searchability, blendability, and findability. On one hand, hiding in a sparsely populated town makes it less likely for someone from that town to be recruited to find the target. But as soon as someone gets recruited, identification becomes trivial. On the other hand, in a city with high density, one might be able to "blend into the crowd." Our results show that, short of *hiding in the middle of nowhere*, one's best bet is to *hide in plain sight*. The role of human mobility in the context of blendability is not completely clear, and certainly warrants further investigation using a more detailed treatment. Models of geographical ties and mobility should explicitly account for variations in density. In particular the deviation from pure rank scaling [50], demonstrating the increased likelihood of city-based users to have longer range ties.

It is worthwhile putting our work in the context of search in social networks. Milgram's landmark "small world" experiment showed that people are, in principle, findable using 6 hops on the global social network [51], a result that has been reaffirmed in the Internet age [52]. However, Milgram searchability relies on people's ability to form a reliable estimate of distance to the target, in order to exploit the large jumps afforded by small world networks [53–55]. For example, if the target is known to be a Professor residing in Kyoto, Japan, one might send it to a friend who lives in Tokyo, Japan, as they are more likely to know someone who lives in Kyoto, who in turn may know someone in academia, and so on. But if information about a target is scarce (e.g. searching for a person in an entire country based only on a mug-shot), we cannot rely on distance estimates. In other words, the problem becomes that of *uninformed* (a.k.a. *blind*) search [56], and thus requires large-scale social mobilization. Having said that, endeavors like the Tag Challenge [57], in which search may benefit from partial knowledge of target location, require elements of both uninformed and *heuristic* search, a topic that deserves further study in the context of social mobilization.

Our work is not without limitations. First, we focused on mobilization processes that are fully driven by social ties. In reality, however, mobilization often also benefits from the use of mass media (e.g. AMBER Alert distributed via radio stations and cable Television) and social media hubs (e.g. highly followed blogs or Twitter accounts). Surely, such media can accelerate social mobilization, as they complement the social diffusion process and seed it over large areas [58, 59]. Another limitation of our work is our use of a simple model of human mobility. For a task like the Network Challenge, this is unlikely to be a problem. However, for scenarios that involve searching for mobile targets, as was the case in the Tag Challenge [57], more sophisticated models of human mobility should be incorporated [33].

Materials and Methods

Materials. High resolution population data was taken from publicly available sources [60] based on US census data [28]. This comprises 7,820,528 cells each with an area of 1km^2, of which 5,060,288 are populated (i.e. 2,760,240 are empty). Empirical data from [6] was used to parameterise the branching factor power law distribution as follows. We exclude the first generation of recruitment directly from the MIT team, as this is anomalously high (164 child recruits) and due to the team's own unique personal association with the task, likely to be atypical. We also exclude 611 single nodes which signed up directly and did not recruit any child nodes, we assume that these are examples of passive recruits which signed up independently. The distribution of the branching factor among a subset of the remaining nodes is described by a power law with mean $\langle R_o \rangle = 0.89$ (See SI Appendix).

Methods.

A set of seed nodes located at MIT is chosen; the number of which matches those initially recruited by the MIT Media Lab team. All of these nodes are active i.e. they continue to recruit themselves in contrast to passive recruits which do not continue the recruitment tree. Each newly activated node looks around in its vicinity (within a distance radius of r_{mob}) and reports any balloon that it sees within that radius. Each newly activated node also chooses an outdegree (a constant number n_{pass} of 'passive' recruits, and a power-law-distributed number n_a of 'active' recruits, where n_{pass} is drawn from the distribution seen in the MIT Red Balloon team's recruitment data). Each chosen friend, passive or active, is chosen to be rank-based with respect to geography, using 1km^2 population density data across the U.S.) with probability 5.5/8, and uniform over population with probability 2.5/8. Each active new recruit selects a delay, chosen from a log-normally distributed waiting time distribution with mean 1.5 days and standard deviation 5.5 days [22] and becomes activated and completes its own recruitment after that time delay.

ACKNOWLEDGMENTS. We thank Wei Pan for assistance with DARPA Balloon Challenge data, Galen Pickard for useful comments, and Mohammed Mekkias for support with HPC resources. Manuel Cebrian acknowledges support from the National Science Foundation under grant 0905645, from DARPA/Lockheed Martin Guard Dog Program under PO 4100149822, and the Army Research Office under Grant W911NF-11-1-0363.

1. Okolloh O (2009) Ushahidi or 'testimony': web 2.0 tools for crowdsourcing crisis information. *Participatory Learning and Action* 59(1):65–70.
2. Hellerstein J, Tennenhouse D (2011) Searching for Jim Gray: a technical overview. *Communications of the ACM* 54:77–87.
3. González-Bailón S, Borge-Holthoefer J, Rivero A, Moreno Y (2011) The dynamics of protest recruitment through an online network. *Scientific reports* 1, 197.
4. Ghonim W (2012) *Revolution 2.0: The Power of the People Is Greater Than the People in Power* (Houghton Mifflin Harcourt).
5. Tang J, *et al.* (2011) Reflecting on the DARPA Red Balloon Challenge. *Communications of the ACM* 54:78–85.
6. Pickard G, *et al.* (2011) Time-critical social mobilization. *Science* 334:509–512.
7. Cebrian M, Coviello L, Vattani A, Voulgaris P (2012) Finding red balloons with split contracts: Robustness to individual's selfishness. *Proceedings of the ACM Symposium on Theory of Computing* pp. 775-788.
8. Naroditskiy V, Rahwan I, Cebrian M, Jennings N (2012) Verification in referral-based crowdsourcing. *PLoS ONE* 7:e45924.
9. (2012) Six degrees of mobilisation. *Economist: Technology Quarterly* .
10. Rahwan I, *et al.* (2013) Global manhunt pushes the limits of social mobilization. *IEEE Computer (In press)* .
11. Golub B, Jackson M (2010) Using selection bias to explain the observed structure of internet diffusions. *Proceedings of National Academiy of Science of the United States of America* 107:10833.
12. Bakshy E, Hofman J, Mason W, Watts D (2011) Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining* (ACM), pp. 65–74.
13. Bond RM, *et al.* (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489:295–298.
14. Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. *Science* 337:337–341.
15. Epstein J (2009) Modeling to contain pandemics. *Nature* 460(7256), 687–687.
16. Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A (2004) Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Physical Review Letters* 92:178701.
17. Wang P, Gonzàlez MC, Hidalgo CA, Barabàsi AL (2009) Understanding the spreading patterns of mobile phone viruses. *Science* 324:1071–1076.
18. Balcan D, *et al.* (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America* 106:21484.
19. Balcan D, *et al.* (2009) Commuting and pandemic prediction. *Proceedings of National Academy of Science of United States of America* 106:21459.
20. Nicolaides C, Felguerso L and Gonzàlez GM (2011) A metric of influential spreading during contagion dynamics through the air transport network. *PLoS ONE* 7:40961.
21. Brockman D, Hufnagel L and Geisel T (2006) The scaling laws of human travel. *Nature* 439:462–465.
22. Iribarren JL, Moro E (2009) Impact of Human Activity Patterns on the Dynamics of Information Diffusion. *Physical Review Letters* 103:038702.
23. Miritello G, Moro E, Lara R (2011) Dynamical strength of social ties in information spreading. *Phys Rev E* 83:045102.
24. Krings J, Calabrese F, Ratti C, Blonde V (2009) Urban gravity: A model for inter-city telecommunication flows. *Journal of Statistical Mechanics.*
25. Ratti C, *et al.* (2010) Redrawing the map of great britain from a network of human interactions. *PLoS ONE* 5:e14248.
26. Barthélemy M (2011) Spatial networks. *Physics Reports* 499:1–101.
27. Liben-Nowell D, Novak J, Kumar R, Raghavan P, Tomkins A (2005) Geographic routing in social networks. *Proceedings of the National Academy of Sciences* 102:11623–11628.
28. Bureau USC (2001), Census 2000: Census Tract Cartographic Boundary Files - U.S. Census Bureau. , [Online; accessed 3-September-2012].
29. Watts D, Dodds P, Newman M (2002) Identity and search in social networks. *Science* 296:1302–1305.
30. Dodds P, Muhamad R, Watts D (2003) An experimental study of search in global social networks. *Science* 301:827–829.
31. Iribarren JL, Moro E (2011) Affinity paths and information diffusion in social networks. *Social networks* 33.2 (2011): 134-142.
32. Ugander J, Karrer B, Backstrom L, Marlow C (2011) The anatomy of the facebook social graph. *arXiv preprint arXiv:1111.4503.*
33. Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility patterns. *Nature* 453:779–782.
34. Cho E, Myers S, Leskovec J (2011) Friendship and mobility: User movement in location-based social networks. *Proceedings of the 17th ACM International Conference on Knowledge Discovery and Data Mining* :1082.
35. Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C (2012) A tale of two cities: Universal patterns in human urban mobility. *PLoS ONE* 7.
36. Bagrow J, Lin Y (2012) Spatiotemporal features of human mobility. *Arxiv preprint arXiv:12020224* .
37. Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C (2012) A tale of many cities: Universal patterns in human urban mobility. *PLoS ONE* 7:e37027.
38. Phithakkitnukoon S, Smoreda Z, Olivier P (2012) Socio-geography of human mobility: A study using longitudinal mobile phone data. *PLoS ONE* 7:e39253.
39. Mack A, Rock I (1998) *Inattentional blindness* (MIT press Cambridge, MA).
40. Most S, *et al.* (2001) How not to be seen: The contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychological Science* 12:9–17.
41. Miller G (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review* 63:81.
42. Milgram S (1970) The experience of living in cities. *Science* 167:1461–1468.
43. Darley J, B Latane (1968) Bystander intervention in emergencies: Diffusion of responsibilities. *Journal of the Learning Sciences* 8:377–383.
44. Lederbogen F, *et al.* (2011) City living and urban upbringing affect neural social stress processing in humans. *Nature* 474:498.
45. Rapoport A, Hawkes R (1970) The perception of urban complexity. *Journal of the American Institute of Planners* 36:106–111.
46. Rosvall M, Trusina A, Minnhagen P, Sneppen K (2005) Networks and cities: An information perspective. *Physical Review Letters* 94:28701.
47. Lee S, Holme P (2012) A greedy-navigator approach to navigable city plans. *Arxiv preprint arXiv:12050537.*
48. Bettencourt L, Lobo J, Helbing D, Kühnert C, West G (2007) Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences* 104:7301.
49. Bornstein M, Bornstein H (1976) The pace of life. *Nature* 259, 557-559.
50. Backstrom L, Sun E, Marlow C (2010) Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World wide web* (ACM, New York, NY, USA), WWW '10, pp. 61–70.
51. Milgram S (1967) The small world problem. *Psychology Today* 61:60–67.
52. Dodds PS, Muhamad R, Watts DJ (2003) An experimental study of search in global social networks. *Science* 301:827–829.
53. Kleinberg J (2000) Navigation in a small world. *Nature* 406(6798), 845–845.
54. Kleinberg J (2000) The small-world phenomenon: an algorithm perspective. *In Proceedings of the thirty-second annual ACM symposium on Theory of computing*(pp. 163–170
55. Adamic LA, Adar E (2005) How to search a social network. *Social Networks* 27:187–203.
56. Knuth D, *et al.* (1973) *Sorting and Searching, The Art of Computer Programming, vol. 3* (Addison-Wesley, Reading, Massachusetts).
57. Firth N (2012) Social media web snares 'criminals'. *New Scientist* 2860:18.
58. Crane R, Sornette D (2008) Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of National Academy of Sciences* 41:15649–15653.
59. Myers SA, Zhu C, Leskovec J (2012) Information diffusion and external influence in networks. *In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* pp. 33-41.
60. Center NNCD (2008), Gridded 1 km Population for the Conterminous United States. [Online; accessed 3-September-2012].

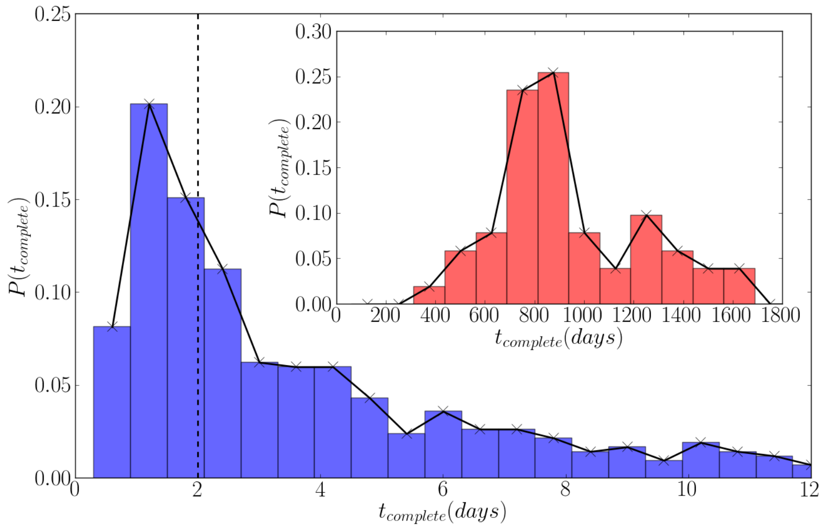

Fig. 1. Histogram of completion times for successful searches out of 500 instances with parameters $n_{pass} = 400$ and $r_{mob} = 2\text{km}$ (blue) and inset for the remaining unsuccessful searches which fail to locate all 10 balloons (red). Dashed vertical line shows completion time of DARPA Network Challenge after MIT team recruitment commenced. The search continues until all agents have acted, due to the heavy tailed waiting time distribution this can take as long as several years. However since the majority of recruits act on much shorter timescales, the searches which succeed in locating all of the balloons drastically truncate this distribution.

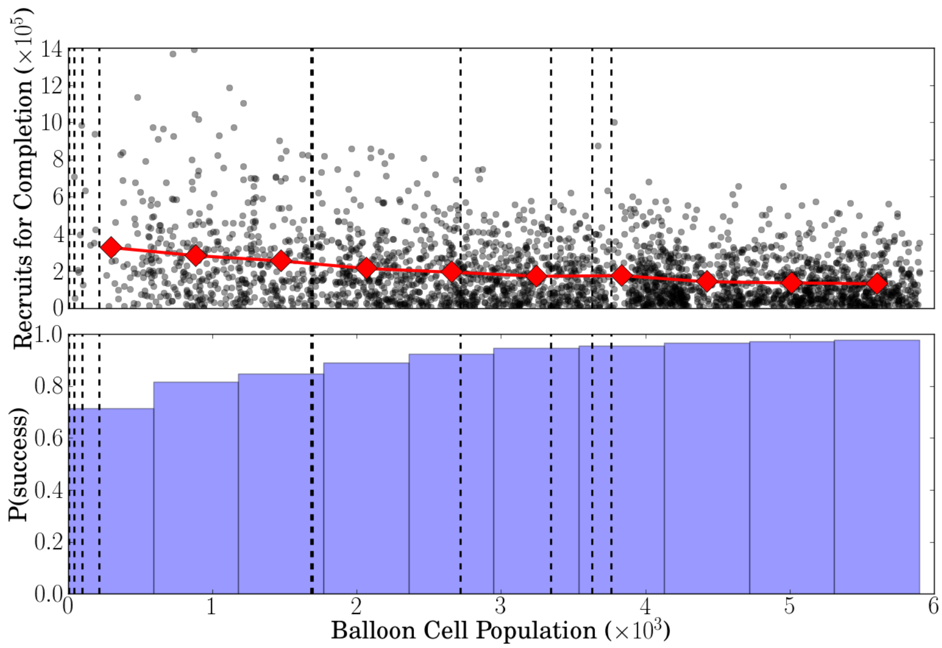

Fig. 2. Scatter plot of number of recruits at completion in a search for a single randomly placed balloon as a function of the population in the cell in which the balloon is placed for 5,000 randomly selected balloon locations. Black dots represent only successful searches (top). Histogram represents the probability to successfully find the balloon. Dashed black vertical lines indicate the populations of the locations used in the DARPA balloon challenge. The red line represents the mean number of recuits for each histogram bin (bottom).

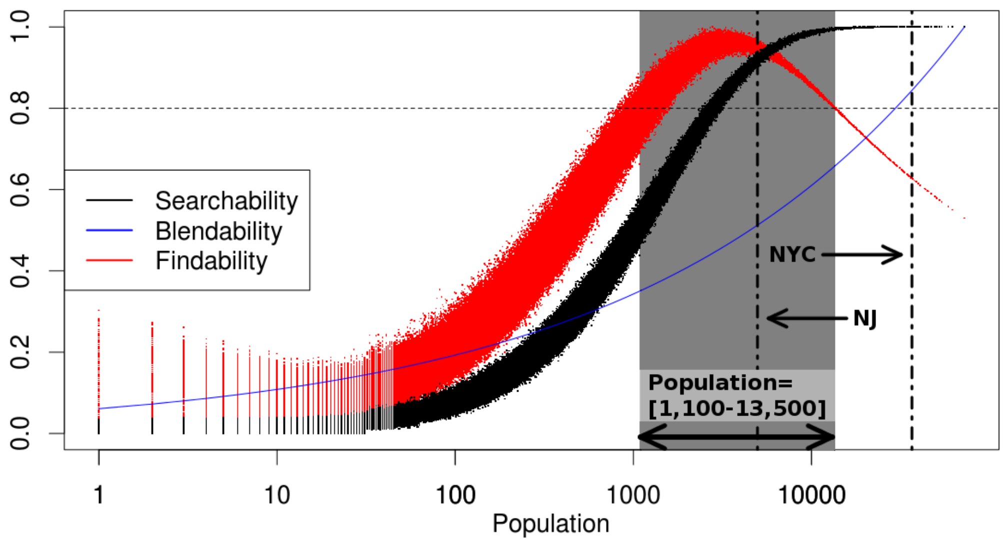

Fig. 3. Scatter plot of searchability (black), population-scaled blendability function (blue) and their ratio, defined as findability (red), as a function of population for all 5060288 cells. The shaded region marks the range of population density for which cells have a findability greater than 0.8. The vertical dashed lines represent Midtown Manhattan, NY and Asbury Park, NJ. [The cells within 15km of the starting cell at MIT have been removed since they are extraordinarily searchable due to their privileged position close to the source of the search, see SI Appendix.]

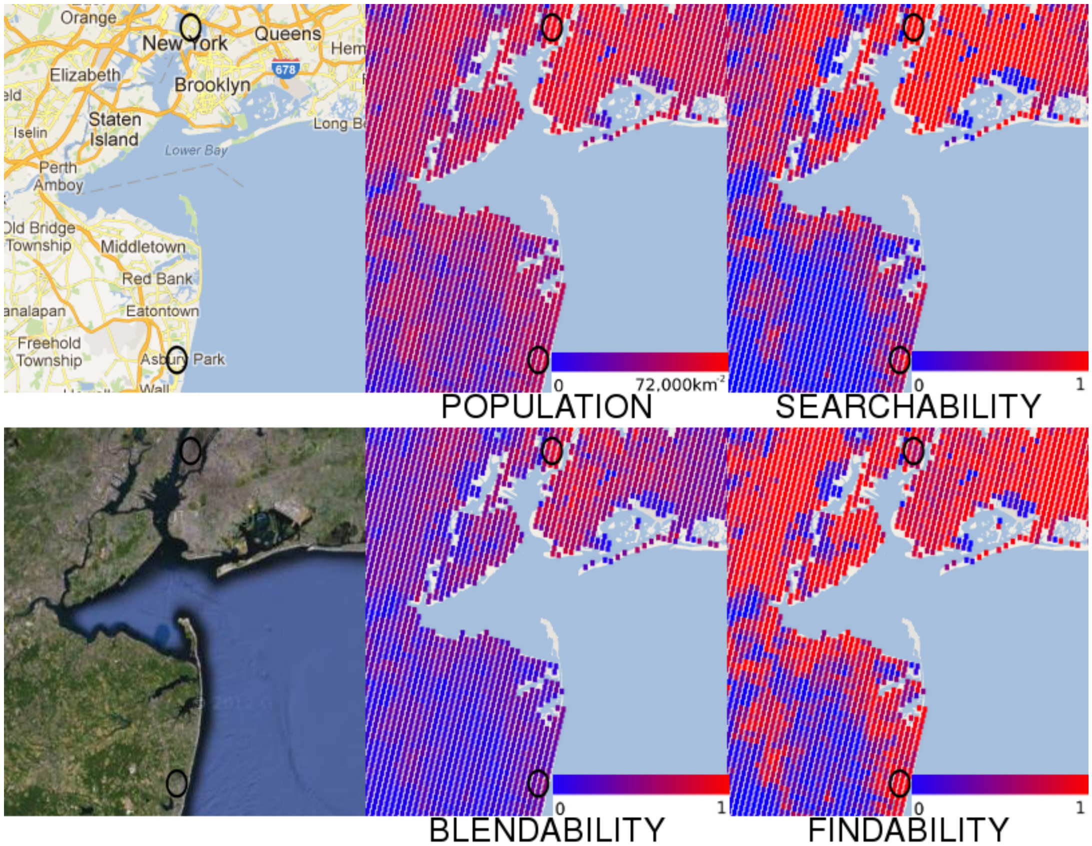

Fig. 4. Map of Manhattan and Asbury Park showing variation of population density, searchability, blendability and findability as well as underlying satellite and road maps. Black circles indicate locations of Manhattan, NY (upper) and Asbury Park, NJ (lower). Population density is on logarithmic scale, all others on linear scale. Unpopulated cells are not shown.