

# The Need for Champions for Approximate Social Search

Yves-Alexandre de Montjoye\*, Jesika Haria\*, Manuel Cebrian\*<sup>†</sup>, and Alex (Sandy) Pentland\*

\*Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139

<sup>†</sup>Department of Computer Science & Engineering, University of California, San Diego, La Jolla, CA 92093

**Abstract**—The world has recently witnessed several disasters, wherein the human cost of the catastrophe depended strongly on the delivery time of critical rescue information, mostly through strangers, on the location and status of individuals. Unfortunately, these are often situations wherein technology ceases to function. Pen-and-paper solutions then become the only viable option for information transfer. Both the propagation of information in networks and social search through traditional digital devices such as mobile phones and e-mails is well-documented in the literature. However, existing models usually fail to take into account the cost of the physical propagation of information. We report on the empirical results of a pen-and-paper system involving time-critical approximate social search. In particular, participants’ goal is to locate and assess the status of other participants without further knowledge regarding their identity. We show that when there is a cost associated with information propagation, the network efficiency is mainly driven by the presence of strongly motivated individuals who actively seek to acquire and propagate information.

## I. INTRODUCTION

The extreme natural or man-made conditions that cause maximum human damage tend to disrupt technological networks during the first hours or days of the aftermath [1]–[3]. This common phenomenon is caused by both the extraneous factors that affect the communication infrastructure [4] and the fragility arising from the inter-connectivity of different technological networks [5]–[7]. Initial rescue operations, which are typically initiated by local residents immediately after the disaster, are thus usually performed in the absence of technological means. These ad-hoc missions whose main focus is the location and assessment of missing relatives, typically involve a coordinated effort for the transmission of information [8]. However, such missions differ from both traditional Social Search [9], [10] and Information Propagation in networks [11], [12] as (1) they involve people who are actively seeking information regarding their relatives, as well as (2) information holders who have no means of knowing the most appropriate way of propagating information, since they lack knowledge about interested receivers; and finally as (3) they use low-tech means for the transmission of time-critical pieces of information. This emphasizes the importance of the study of low-tech transmission of information for the success of post-disaster rescue operations.

Social search is indeed one of the seminal problems of modern social science [9]. The distributed routing of information from senders to unknown recipients, popularly known as the

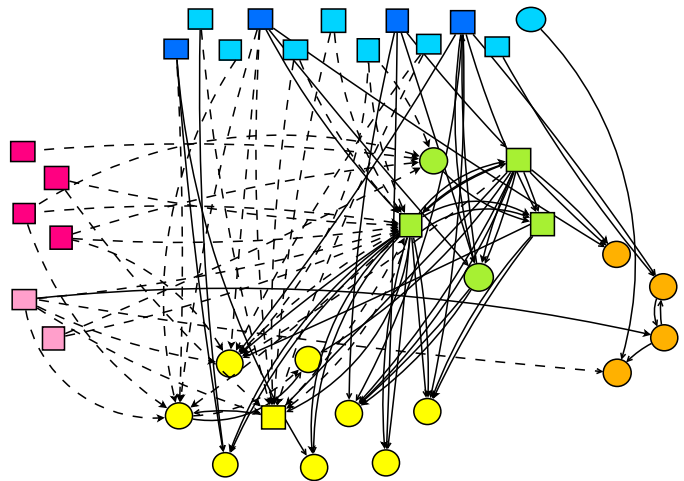


Fig. 1. The visible part of the network underlying the propagation of information in our social search system. As described in Table II, Pink are sources, Exceptions in Blue, Propagators in Green and Sink in Yellow/Orange

small world problem, has illustrated the complexity of the approach for more than 50 years [10], [13]–[15]. It has been empirically demonstrated that messages are able to reach the destination in a small number of routing steps [16] — the so-called “small world” effect. However, it has been observed that most chains fail to reach the destination after a few steps. This high attrition rate is prevalent across multiple experiments using not only traditional mail but also email where the cost of information transfer is effectively zero [10]. This has generally been attributed to the lack of individual incentive amongst participants. In the absence of technological networks, time-critical social search poses two new additional challenges. On the one hand, the task is time-critical, as relief teams usually arrive after 24 hours whilst survival rates dwindle rapidly with increasing time elapsed, whereas the small world problem experiments unfold on the scale of months. On the other hand, the transmission of information has a physical cost associated with mobility and accessibility, as there are no communication infrastructures, mail relays, or digital communication networks. This significantly increases the cost of information propagation in low-tech social networks as compared to its technological counterpart.

The important behavioral specifics of search operations is

Target	Status	Who told you? (id)
22	Green / Orange / Red	.....
39	Green / Orange / Red	.....
57	Green / Orange / Red	.....
59	Green / Orange / Red	.....
<b>62</b>	<b>Green / Orange / Red</b>	.....
69	Green / Orange / Red	.....
<b>73</b>	<b>Green / Orange / Red</b>	.....

Fig. 2. Sample paper form received by a participant. The bold rows represent the targets for which the participant gets rewarded in case of finding the true status.

a subject that can hardly be studied in a laboratory setting [17]. We have performed a preliminary empirical investigation into time-critical transmission of information in the absence of a technological network. In this paper we describe the test setting and analyze the data generated during its course. We show that, in this short time-span, when there is a cost associated with the propagation of information, the efficiency of the network is mainly driven by the presence of strongly motivated individuals that actively seek to acquire and propagate information.

## II. SETUP

The key determinants of the efficiency of an approximate social search were analyzed as part of a demonstration at the MIT Media Laboratory. The test was voluntarily interspersed with many other activities for a period of one day. In order to better simulate a real disaster, the participants were given minimal information and constraints. Approximately 10% of the participants were randomly assigned to be target, symbolizing people that might need assistance in a disaster situation. These target were also randomly assigned a color (“green”, “orange” or “red”) which we call status and which correspond to increasingly high-risk individual. Each participant was randomly assigned 5 people (e.g. “relatives”) whose status was of interest to them. It is important to notice that while each participant’s goal was to merely find the status of these individuals, the pen-and-paper solutions (Fig. 2) were designed such that the participant could record and report the status of all targets as well as the identification number of the individual who gave them the information. They were rewarded following a winner-takes-all model for the information they were able to collect regarding their relatives.

Most participants came in with strong connections with participants from their same corporation, forming cliques of small size. In addition, they could also have had a few weak connections with people they met at previous events. We can therefore hypothesize that the underlying network is likely to be sparse and modular. A posteriori, we were able to observe that while most information was transferred through traditional low-tech means such as word of mouth, there were more innovative means employed as well: white-board messages, as well as high-frequency locations of transit such as the coffee machine were also used. All data regarding propagation was reported by the participants themselves. Cases wherein

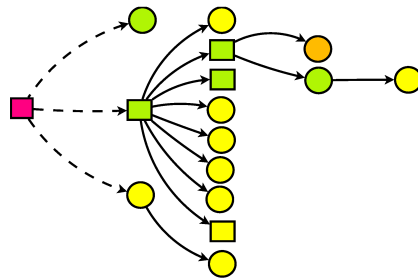


Fig. 3. Sample tracing of a piece of information. The target is the root of the propagation tree, and the participants that did not propagate the information are leaves.

the information was received but no record of its source was provided, such as messages propagated through board messages and the like, were reported as having an indirect link to the original target.

## III. ROLES IN INFORMATION PROPAGATION

Previous studies have emphasized the influence of individual participants characteristics for social search and information propagation [10], [15]. In order to understand the role played by each participant in the propagation of information, we defined the following coding scheme (Table I). For example, PRD represents nodes to propagate right information directly while the RWD are nodes that were to receive wrong information directly. If we assume the propagation of specific information to be directional, each individual can either propagate or receive information. We can further suppose that low-tech systems are intrinsically more prone to errors and bad retransmission than digital ones. Approximately 32% of the information transmitted was found to be incorrect and were thus labelled as wrong. As discussed before, we chose to not constrain the system on how information should be transmitted and we therefore observed that indirect low-tech means were successfully employed.








TABLE I  
PROFILES

Propagate	<b>(P)</b>	Receive	<b>(R)</b>
Right	<b>(R)</b>	Wrong	<b>(W)</b>
Direct	<b>(D)</b>	Inferred	<b>(I)</b>

We restricted each node to a maximum of three dominant profiles expressed out of the  $2^3$  possible combination. We then clustered each node in the space defined by their profile (averaged per node). A personality is thus a cluster of profiles that are expressed by the same nodes. As we can see in Table. 2, the clusters are mostly distinct and personalities do not overlap.

The dominant and clandestine sources represent the targets or the main sources of information (see Fig. 1). Clandestine sources belong to “fringe” networks that are typically narrow and do not infect too many nodes. On the other hand, dominant sources have wider networks and more active nodes

TABLE II  
PERSONALITIES

	Color	Dominant profiles	Percentage	Characteristic features
Made-up exceptions		PWI	11.10%	Propagate only wrong information indirectly.
Semi-made-up exceptions		PWI, PRI	5.60%	Propagate both wrong and some right information but always indirectly.
Clandestine Sources		PRD, PRI	16.70%	Do not receive any information. Propagate only right information mostly indirectly which is why we termed their actions as clandestine. They typically do not have a very high degree.
Dominant Sources		PRD, RRD	13.90%	More actively involved in the network; while they receive information their main activity is to propagate right information. They typically have a higher degree than the clandestine sources.
Main propagators		PRD, PWD,RRD, RWD	13.90%	Actively involved in network as major contributors to both receiving and propagating information. This put them at the core of the visible underlying network.
Large sinks		RRD, RWD	22.20%	At the very bottom of the information network, they mostly receive both right and wrong information but do not pass it along.
Small sinks		RRD, PRD	16.70%	Similar to the large sink, they typically get information from less than 2 different sources but pass right information to few people.

that not only propagate but also receive information. The main propagators both receive and propagate data from a variety of different sources. These nodes are likely to have a greater average and maximum branching factor, although depth remains consistent with the dominant sources. Finally, the network consists of small and large sinks that receive information but barely pass it on. In social search terms, these are the nodes that are responsible for the high attrition rate. Large sinks mainly receive information from more than two nodes but do not pass it onto other sources while small sinks gather the information from nodes with clandestine sources, acting as funnels for these narrow networks.

#### IV. THE CHAMPIONS

Beyond the visible network, we also retrace the path taken by each piece of information. The propagation of the status of a target is defined by a tree with the target as them root and the sinks as leaves. An example of such tree can be seen in Fig. 3. This allows us to analyze the main structural determinants for efficient information propagation measured as the size/number of nodes of the tree.

In this analysis, we posited that each participant who received the information was as likely to propagate it as the target was even though the time needed for the information to reach a participant at a given depth certainly slightly influenced the number of participants who could have received the information from him or his children. Furthermore, we interestingly found no significant difference in the propagation of right and wrong information. While analyzing trees with either a majority of right or wrong information, neither the size of the tree  $s$  ( $p = 0.32$  with a 2-sided t-test) nor its depth ( $d$ ) ( $p = 0.14$  with a Chi-squared test for categorical variables) were statistically significant.

For each tree, we can then compare its size to its depth and branching factor. It is important to notice that both depth and

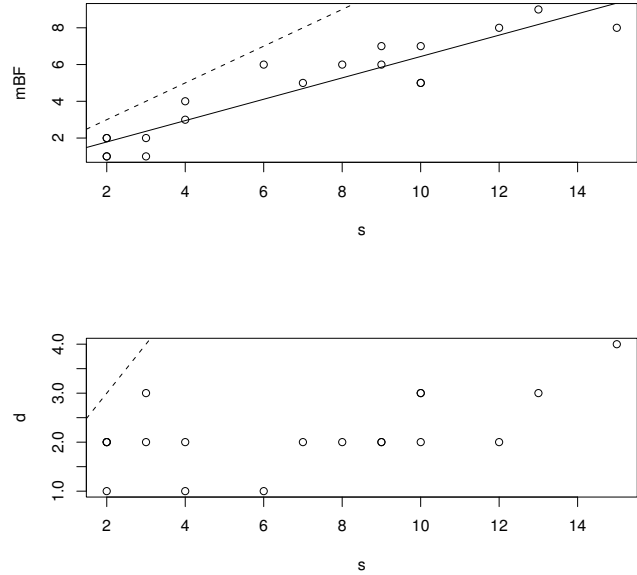


Fig. 4. Correlation between the size of the tree ( $s$ ), its depth ( $d$ ) and its max branching factor (mBF). The dashed lines represent the baseline to which we compare for correlation ( $\rho = 1$ ).

branching factor set a minimum on the size of the tree. Thus by definition, no point can be placed strictly above the curves  $s = d$  and  $s = \text{mBF}$ . We correct for this effect while regressing and show the base line in Fig. 4. We observe a strong correlation between its corrected size and its maximum branching factor (mBF) ( $\rho = 0.6281, p = 0.003$ , Fig. 4 Above). We find no significant correlation between its corrected size and depth ( $p > 0.05$ , Fig. 4 Below). The correlation with other metrics

related to the branching factor (2-max, mean,...) is significantly smaller than that with the maximum branching factor.

## V. DISCUSSIONS

These results present interesting similarities to a previous study [18] that showed the difficulty in predicting the ability of Twitter users to trigger cascades. Furthermore, they indicate that the efficiency of the diffusion of information in a low-tech social network is typically dependent on the presence of a single highly motivated individual. Dodds *et al.* already showed that even small variations in participation rates can have far-reaching impacts on network penetration and target reachability, similarly emphasizing the importance of personal incentive in successful social search [10].

Standard information propagation in a network is usually considered to occur through a simple contagion process. A single contact with the information is usually sufficient to be “infected.” Once infected, the probability of this individual to infect another is fixed, potentially due to the low cost of transmitting information. By contrast, the complex contagion process, where contact with multiple infected individuals is needed for successful infection, is believed to be essential for triggering action or behavioral changes [19]. In [20], Centola and Macy suggest different scenarios where complex contagion might be needed such as “credibility” (i.e. when an individual waits to receive the information from multiple sources before acting on it) or “legitimacy” (i.e. the reluctance of an individual to take the risk of being part of a movement before it reaches a critical threshold). In the specific setting of approximate social search, we speculate that the high cost of information transfer might have prompted a shift from a simple to a complex contagion process for information transmission. Furthermore, as each participant had a mixture of weak and strong ties, a highly motivated individual could then be someone who is proactive enough and willing to invest a lot of energy into triadic closure (as suggested in [21]) with their acquaintance connections thereby allowing for complex contagion to occur.

Once again, limited experimental evidence points towards the importance of personal incentives in successful approximate social search. We can conjecture that in a disaster situation, the higher stakes could induce more people to take a more proactive role in information transfer, thereby activating “latent links” in their networks allowing for complex contagion to happen [22]. More in-depth experiments are needed to further assess the generality of this empirical observations. Our ultimate goal is to understand the reach of personalities and their influence in social search. Given a larger sample, we would like for example to better quantify how many highly motivated individuals would be needed to reach  $N$  missing individuals. It would also be interesting to run a controlled experiment assessing whether certain personalities affect network flow, and if so, in what manner. These are unanswered questions that provide the basis for further research in this area.

## ACKNOWLEDGMENT

The authors would like to thank Ben Waber and Riley Crane for their help while deploying the analyzed system. Yves-Alexandre de Montjoye thanks the Belgian American Educational Foundation (B.A.E.F.) for its financial support. Manuel Cebrian is supported by the National Science Foundation Grant No. 0905645.

## REFERENCES

- [1] E. Brattberg and B. Sundelius, “Mobilizing for international disaster relief: Comparing us and eu approaches to the 2010 haiti earthquake,” *Journal of Homeland Security and Emergency Management*, vol. 8, no. 1, p. 24, 2011.
- [2] F. Yamazaki, C. Zavala, S. Nakai, S. Koshimura, T. Saito, and S. Midorikawa, “Enhancement of earthquake and tsunami disaster mitigation technology in peru: A satreps project,” in *Proceedings of the 7th International Conference on Urban Earthquake Engineering*, 2010, pp. 1501–1506.
- [3] C. Butts, M. Petrescu-Prahova, and B. Cross, “Responder communication networks in the world trade center disaster: Implications for modeling of communication within emergency settings,” *The Journal of mathematical sociology*, vol. 31, no. 2, pp. 121–147, 2007.
- [4] M. Battaini and S. Dyke, “Fault tolerant structural control systems for civil engineering applications,” *Journal of Structural Control*, vol. 5, no. 1, pp. 1–26, 1998.
- [5] S. Buldyrev, R. Parshani, G. Paul, H. Stanley, and S. Havlin, “Catastrophic cascade of failures in interdependent networks,” *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.
- [6] A. Vespignani, “The fragility of interdependency,” *Nature*, vol. 464, no. 7291, p. 15, 2010.
- [7] X. Huang, J. Gao, S. Buldyrev, S. Havlin, and H. Stanley, “Robustness of interdependent networks under targeted attack,” *Physical Review E*, vol. 83, no. 6, p. 065101, 2011.
- [8] G. Kovács and K. Spens, “Humanitarian logistics in disaster relief operations,” *International Journal of Physical Distribution & Logistics Management*, vol. 37, no. 2, pp. 99–114, 2007.
- [9] S. Milgram, “The small world problem,” *Psychology today*, vol. 2, no. 1, pp. 60–67, 1967.
- [10] P. Dodds, R. Muhamad, and D. Watts, “An experimental study of search in global social networks,” *Science*, vol. 301, no. 5634, p. 827, 2003.
- [11] J. Iribarren and E. Moro, “Impact of human activity patterns on the dynamics of information diffusion,” *Physical review letters*, vol. 103, no. 3, p. 38702, 2009.
- [12] V. Preciado and A. Jadbabaie, “Moment-based analysis of spreading processes from network structural information,” *Arxiv preprint arXiv:1011.4324*, 2010.
- [13] J. Travers and S. Milgram, “An experimental study of the small world problem,” *Sociometry*, vol. 32, no. 4, pp. 425–443, 1969.
- [14] S. Schettler, “A structured overview of 50 years of small-world research,” *Social Networks*, vol. 31, no. 3, pp. 165–178, 2009.
- [15] D. Watts, P. Dodds, and M. Newman, “Identity and search in social networks,” *Science*, vol. 296, no. 5571, p. 1302, 2002.
- [16] J. Kleinberg, “Navigation in a small world,” *Nature*, vol. 406, no. 6798, pp. 845–845, 2000.
- [17] M. Zelditch Jr, “Can you really study an army in the laboratory,” *A sociological reader on complex organizations*, pp. 528–539, 1969.
- [18] E. Bakshy, J. Hofman, W. Mason, and D. Watts, “Everyone’s an influencer: Quantifying influence on twitter,” in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011, pp. 65–74.
- [19] D. Centola, “The spread of behavior in an online social network experiment,” *science*, vol. 329, no. 5996, p. 1194, 2010.
- [20] D. Centola and M. Macy, “Complex contagions and the weakness of long ties,” *American Journal of Sociology*, vol. 113, no. 3, p. 702, 2007.
- [21] M. Granovetter, “The strength of weak ties,” *The American journal of sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.
- [22] N. Hassanpour, “Media disruption exacerbates revolutionary unrest: Evidence from mubaraks natural experiment,” 2011.