Time-Critical Social Mobilization
Galen Pickard *et al.*
Science **334**, 509 (2011);
DOI: 10.1126/science.1205869

18. G. Zhang, J. Zhang, S. Liu, *J. Atmos. Chem.* **57**, 41 (2007).
19. G. Onitsuka, I. Uno, T. Yanagi, J.-H. Yoon, *J. Oceanogr.* **65**, 433 (2009).
20. J. Kang, B. C. Cho, C. B. Lee, *Sci. Total Environ.* **408**, 2369 (2010).
21. Seawater N* values in the euphotic layer were affected by the supply of N from the upper thermocline (formed by remineralization of organic matter), which occurs on time scales less than 1 year; this is in contrast to the monthly time scale on which the atmospheric nitrogen supply into the euphotic layer was measured. The mismatch in time scales of these two processes (N supply from below versus air-N^{ANTH} deposition) could affect the correlations between the values of surface N* and air-N^{ANTH} deposition. Therefore, to remove seasonal fluctuations and to highlight the interannual or long-term trends, the data for air-N^{ANTH} deposition and seawater N* were smoothed using a 2-year moving average before undertaking comparisons. This data treatment minimized potential biases in the correlations in Fig. 3, enabling comparison of the two parameters.
22. P. G. Falkowski, R. T. Barber, V. Smetacek, *Science* **281**, 200 (1998).
23. Y. W. Watanabe, M. Shigemitsu, K. Tadokoro, *Geophys. Res. Lett.* **35**, L01602 (2008).
24. The riverine influence is largely confined to Chinese coastal waters during winter (October to April), whereas, in summer (May to September), the freshwater plume extends northeast toward Cheju Island (on which an air-monitoring station, marked "A" in Fig. 3E, is located) (*30*). However, the impact of the river plume rapidly diminished away from the source, as indicated by a rapid decrease in N concentration from 40 μM at the mouth of the Changjiang River to 3 μM in offshore waters, ~300 km distant from the river mouth (Fig. 3F) (*31*). About 75% of the total nitrogen load from the Changjiang River is added to the East China Sea during summer, with the remainder added during winter (*32*). In addition, we found no statistically significant correlations between seawater N* values in boxes 6 and 7 (wherein the effect of the Han River is likely strong) and N* values measured over the past 15 years in waters near the mouth of the Han River (fig. S6), indicating a negligible contribution from the Han River nitrogen flux into the Yellow Sea.
25. Contributions from N_2 fixation are unlikely because oceanic conditions in our study area did not favor N_2 fixation. With rare exceptions (*33*, *34*), most N_2 fixation blooms occur in tropical oceans in which the fixed N concentration is nearly zero and the sea surface temperature is generally higher than 25°C (*35*, *36*). Most of our study areas featured sea surface temperatures below 25°C, except during the summer season, and the surface N concentrations were generally greater than the detection limits.
26. S. C. Doney, *Science* **328**, 1512 (2010).
27. C. A. Klausmeier, E. Litchman, T. Daufresne, S. A. Levin, *Nature* **429**, 171 (2004).
28. K. R. Arrigo, *Nature* **437**, 349 (2005).
29. T. S. Weber, C. Deutsch, *Nature* **467**, 550 (2010).
30. H. Ichikawa, R. C. Beardsley, *J. Oceanogr.* **58**, 77 (2002).
31. B. D. Wang, X. L. Wang, R. Zhan, *Estuar. Coast. Shelf Sci.* **58**, 127 (2003).
32. Z. Dai, J. Du, X. Zhang, N. Su, J. Li, *Environ. Sci. Technol.* **45**, 223 (2011).
33. G. H. Park *et al.*, *Limnol. Oceanogr.* **53**, 1697 (2008).
34. P. H. Moisander *et al.*, *Science* **327**, 1512 (2010).
35. D. Karl *et al.*, *Biogeochemistry* **57-58**, 47 (2002).
36. C. S. Davis, D. J. McGillicuddy Jr., *Science* **312**, 1517 (2006).
37. H.-J. Lie, C.-H. Cho, J.-H. Lee, S. Lee, *J. Geophys. Res.* **108**, (C3), 3077 (2003).

Acknowledgments: This work would not have been possible without the efforts of many scientists who contributed data to National Fisheries Research and Development Institute and Japan Meteorological Agency for the seawater nutrient data and to Acid Deposition Monitoring Network in East Asia and Korea Meteorological Administration (Climate Change Information Center) for the atmospheric deposition data. The preparation of the manuscript was supported by the Mid-career Researcher Program (2009-0084756) of the Korea National Research Foundation (K.L.). Partial support was provided by the Korea Meteorological Administration Research and Development Program under grant RACS_2010-1006 (K.L.). Support for H.J.J. was provided by the Ministry of Land, Transport, and Maritime Affairs' Ecological Disturbance Program. Support for R.G.N. was provided by NASA's Ocean Biology and Biogeochemistry Program (contract NNX08AO25G).

Supporting Online Material
www.sciencemag.org/cgi/content/full/science.1206583/DC1
Materials and Methods
SOM Text
Figs. S1 to S6
Tables S1 and S2
References

5 April 2011; accepted 29 August 2011
Published online 22 September 2011;
10.1126/science.1206583

Time-Critical Social Mobilization

Galen Pickard,[1,2]* Wei Pan,[1]* Iyad Rahwan,[1,3]* Manuel Cebrian,[1]* Riley Crane,[1] Anmol Madan,[1] Alex Pentland[1]†

The World Wide Web is commonly seen as a platform that can harness the collective abilities of large numbers of people to accomplish tasks with unprecedented speed, accuracy, and scale. To explore the Web's ability for social mobilization, the Defense Advanced Research Projects Agency (DARPA) held the DARPA Network Challenge, in which competing teams were asked to locate 10 red weather balloons placed at locations around the continental United States. Using a recursive incentive mechanism that both spread information about the task and incentivized individuals to act, our team was able to find all 10 balloons in less than 9 hours, thus winning the Challenge. We analyzed the theoretical and practical properties of this mechanism and compared it with other approaches.

In crowdsourcing, an interested party provides incentives for large groups of people to contribute to the completion of a task (*1*, *2*). The nature of the tasks and the incentives vary substantially, ranging from monetary rewards, to entertainment, to social recognition (*3*–*7*).

A particularly challenging class of crowdsourcing problems requires not only the recruitment of a very large number of participants but also extremely fast execution. Tasks that require this kind of time-critical social mobilization include search-and-rescue operations, hunting down outlaws on the run, reacting to health threats that need instant attention, and rallying supporters of a political cause.

To mobilize society, one may turn to mass media. However, the ability to use mass media can be inhibited for many reasons, such as telecommunications infrastructure breakdown. In such cases, one must resort to distributed modes of communication for information diffusion. For example, in the aftermath of Hurricane Katrina amateur radio volunteers helped relay 911 traffic for emergency dispatch services in areas with severe communication infrastructure damage (*8*). At other times, the nature of the task itself necessitates socially driven diffusion because it requires tight involvement that can only be generated socially.

Another common characteristic of these social mobilization problems is the presence of some sort of search process. For example, search may be conducted by members of the mobilized community for survivors after a natural disaster. Another kind of search attempts to identify individuals within the social network itself, such as finding a medical specialist to assist with a challenging injury.

There is growing literature on search in social networks. It has long been established that social networks are very effective at finding target individuals through short paths (*9*), and various explanations of this phenomenon have been given (*10*–*13*). However, the success of search in social mobilization requires individuals to be motivated to actually conduct the search and participate in the information diffusion; indeed, the majority of chains observed empirically terminate prematurely. Providing appropriate incentives is a key challenge in social mobilization (*14*, *15*).

Recognizing the difficulty of time-critical social mobilization, the Defense Advanced Research Projects Agency (DARPA) announced the DARPA Network Challenge in October 2009. Through this challenge, DARPA aimed to "explore the roles the Internet and social networking play in the timely communication, wide-area team-building, and urgent mobilization required to solve broad-scope, time-critical problems" (*16*). The challenge was to provide coordinates of 10 red weather balloons placed at different locations in the continental United States. According to DARPA, "a senior analyst at the National Geospatial-Intelligence Agency characterized the problem as impossible" by conventional intelligence-gathering methods (*17*).

We, as the Massachusetts Institute of Technology (MIT) team, won the challenge (*18*), completing the task in 8 hours and 52 min. In ~36 hours before the beginning of the challenge, we were able to recruit almost 4400 individuals through a recursive incentive mechanism. Between

[1]Media Laboratory, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [2]Lincoln Laboratory, MIT, Lexington, MA 02420, USA. [3]Computing and Information Science Program, Masdar Institute of Science and Technology, Masdar City, Abu Dhabi 54224, United Arab Emirates.

*These authors contributed equally to the paper.
†To whom correspondence should be addressed. E-mail: pentland@mit.edu

50 and 100 other teams participated in the DARPA Network Challenge (*17*). Although no other team located the 10 balloons, The Georgia Institute of Technology (GaTech) team placed second by locating nine balloons within 9 hours. Two more teams found eight balloons (DudeItsaBalloon and Rodriguez-Chang), and five other teams found seven balloons. Variations in the strategies of the competing teams reflected differences in how social media can be tailored in order to fit a given task (*19*).

The MIT team's strategy for public collaboration was to use the $40,000 prize money that would be awarded to the winning team as a financial incentive structure rewarding not only the people who correctly located balloons but also those connecting the finder to us. Should we win, we would allocate $4000 in prize money to each of the 10 balloons. We promised $2000 per balloon to the first person to send in the correct balloon coordinates. We promised $1000 to the person who invited that balloon finder onto the team, $500 to whoever invited the inviter, $250 to whoever invited that person, and so on. The underlying structure of the "recursive incentive" was that whenever a person received prize money for any reason, the person who invited them would also receive money equal to half that awarded to their invitee (fig. S1).

Our approach ("mechanism") was based on the idea that achieving large-scale mobilization requires incentives at the individual level to execute the task as well as to be actively involved in the further recruitment of other individuals through their social networks. A formal model of the approach is in the supporting online material (SOM) text. In this diffusion-based task environment, agents become aware of tasks as a result of either (i) being directly informed by the mechanism through advertising or (ii) being informed through recruitment by an acquaintance agent (*20*).

Our approach can be seen as a variant of the Query Incentive Network model of Kleinberg and Raghavan (*21*), in which a query propagates over a network through a subcontracting process, and the answer propagates back once it is found (SOM text). The use of incentives to spread information on a social network is also frequent in referral marketing programs, which encourage existing customers to promote the product among their peers—for example, by giving the customer a coupon for each friend recruited (*22*). A fundamental difference between these techniques and ours is that our reward scales with the size of the entire recruitment tree (because larger trees are more likely to succeed), rather than depending solely on the immediate recruited friends.

Our mechanism's performance compares well with previous studies on search and recruitment in social networks. One measure of success is the size of the cascades, both in terms of number of nodes, as well as depth. In a study of the spread of online newsletter subscriptions (*23*), in which individuals were rewarded for recommending the newsletter to their friends, the 7188 cascades varied in size between 2 and 146 nodes, with a maximum depth of eight steps, over a time span of 3 months. In our data, if we ignore the MIT root node there were 845 trees recruited within 3 days. Examples of these trees are shown in fig. S5. The largest tree contained 602 nodes, and the deepest tree was 14 levels deep. The distribution of tree/cascade depth is shown in Fig. 1A. Furthermore, a power-law distribution of tree/cascade size with exponent −1.96, as predicted by models of information avalanches on sparse networks, is shown in Fig. 1B (*24*).

Previous empirical studies reported attrition rates, which measure the percentage of nodes that terminate the diffusion process. For example, in a study of e-mail–based global search for 18 target persons, attrition rate varied between 60 and 68% in 17 out of the 18 searches performed (*15*). In another study of the diffusion of online recommendations, an attrition rate of 91.21% was reported, despite providing incentives to participants by offering them a chance in a lottery (*24*). In the DARPA Network Challenge, if we ignore isolated single nodes our mechanism achieved a significantly lower attrition rate of 56%.

Another measure of performance for social mobilization processes is the branching factor (also known as the reproductive number), which

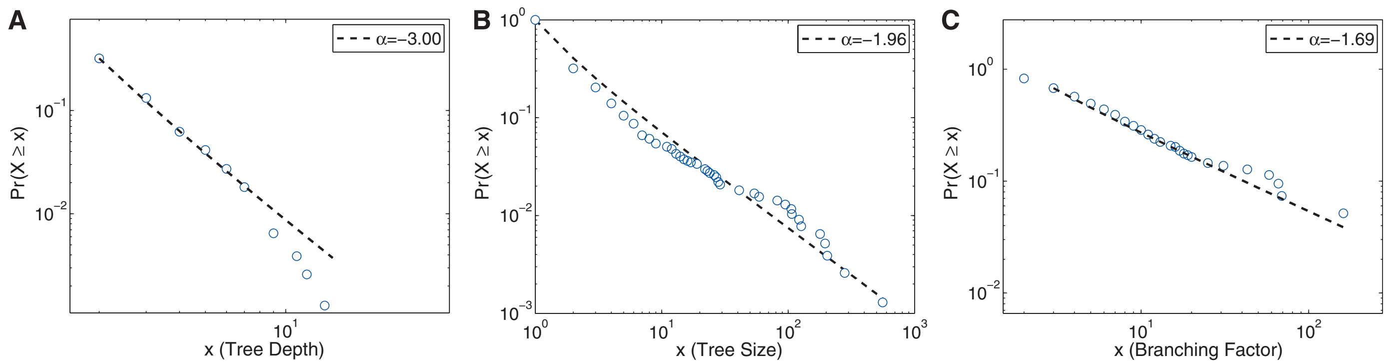


Fig. 1. (**A**) Distribution of tree depth on a log-log scale with a power law fit. (**B**) Distribution of tree size on a log-log scale with a power law fit. (**C**) Distribution of the branching factor on a log-log scale with a power law fit.

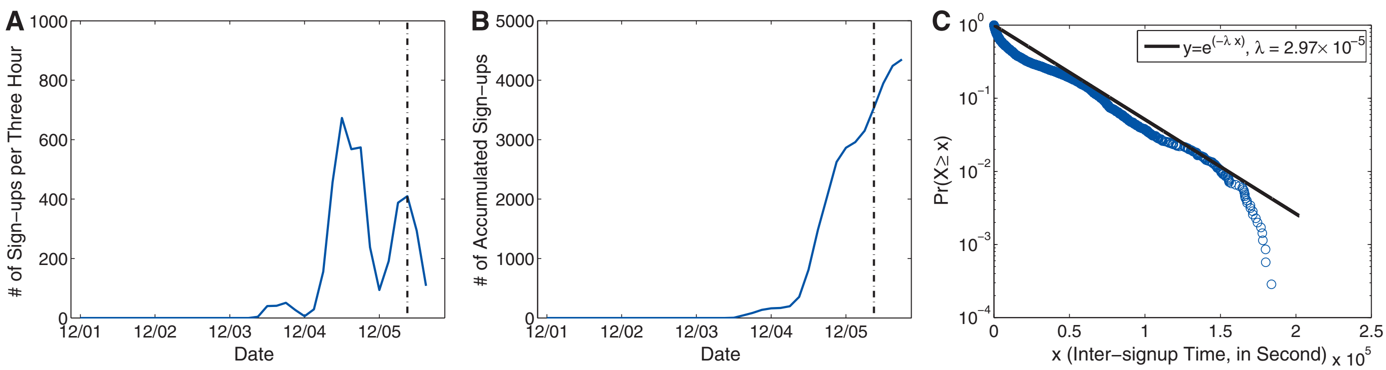


Fig. 2. (**A**) Number of people recruited over time up to the winner announcement. The dotted line indicates the time the balloons were launched into their positions by DARPA. (**B**) Cumulative number of people recruited over time. (**C**) Complementary cumulative distribution of the inter–sign-up time on a semi-log scale with an exponential fit. Shown is a larger-than-exponential drop off at the end of the graph, wich is due to the time-critical nature of the task.

is the number of people recruited by each individual. Previous empirical studies reported diverse, though mostly low, observations. In a study of the spread of support for online petitions, dissemination was very narrow, with >90% of nodes having exactly one child (*25*), which others have attributed to a selection bias, observing only large diffusions (*26*). In our data, the average branching factor was 0.93 if we exclude single-node trees (0.80 if we include single-node trees). The branching factor follows a power-law distribution, suggesting that certain individuals played an important role in dissemination by recruiting a very large number of people (Fig. 1C). Our data also compares very favorably with the newsletter subscription experiment mentioned above, in which spreaders invited an average of 2.96 individuals but were only able to cause 0.26 individuals to sign up on average (*23*). More generally, our data indicates that the branching factor appears to be closer to the tipping point (branching factor of 1), above which large cascades ensue. However, the cascade was finite because of the completion of the task.

The dynamics of recruitment over time are shown in Fig. 2, A and B, highlighting two bursts of day-time recruitment activities on Friday and Saturday just before DARPA launched the balloons into their locations. In contrast with the newsletter subscription experiment (*23*), in which diffusion experienced a continuous decay, these bursts enabled our mechanism to amass a large number of people quickly.

Moreover, in the newsletter subscription experiment the dynamics of diffusion were slow, which was attributed to a heterogeneous, non-Poisson distribution of individuals' response time. We observed an exponential distribution of inter–sign-up time (Fig. 2C) (*27*). This contrasts with the empirically observed power-law distribution of inter-response time in human activity (*28*, *29*) and information cascades (*23*).

In message-routing tasks, it has been argued that the ability of individuals to find a target with an approximately known location is largely attributed to people's ability to exploit geography (*11*, *15*). To investigate this, we plotted the probabilistic density distribution of distances between two parties in a successful recruitment (Fig. 3). We compared this data with a best-fit model that explains the distribution of friendship over geographical distance in the popular LiveJournal online community (Fig. 3) (*30*). Our data exhibited higher likelihood of distant connections compared with the model by Liben-Nowell *et al.* (*30*). Furthermore, this was confirmed by applying the Kolmogorov-Smirnov test, comparing our data with random samples drawn from the model ($P < 10^{-100}$). This suggests that people may have exerted greater effort in recruiting distant friends. This might be due to an expectation that increasing the geographic spread of their recruitment sub-tree is likely to increase their expected reward.

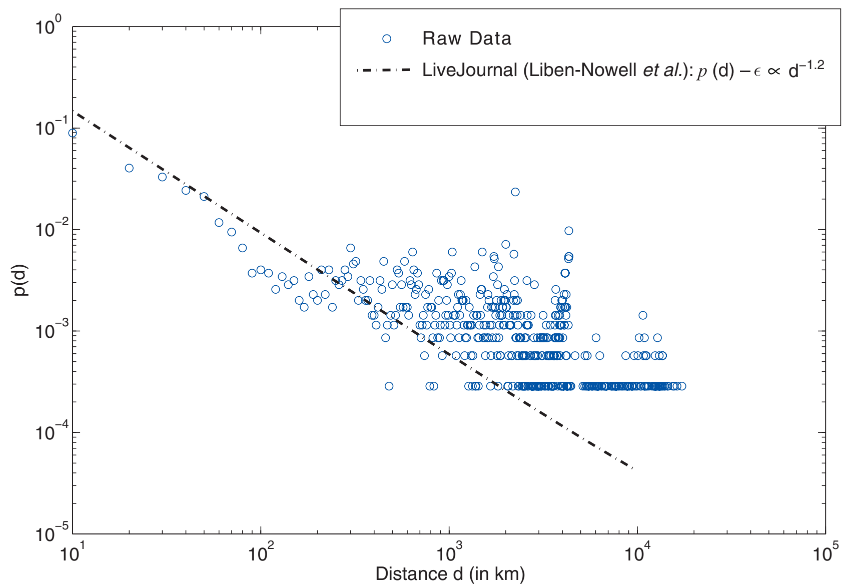


Fig. 3. Distribution of distance between recruiter nodes and their recruits. The dotted line shows the best-fit rank-based friendship model by Liben-Nowell *et al.* (*30*). We apply the same treatment to our data points as in Liben-Nowell *et al.* by rounding distances to multiples of 10 km. Approximate geographic locations were discovered from users' Internet provider addresses during sign-up.

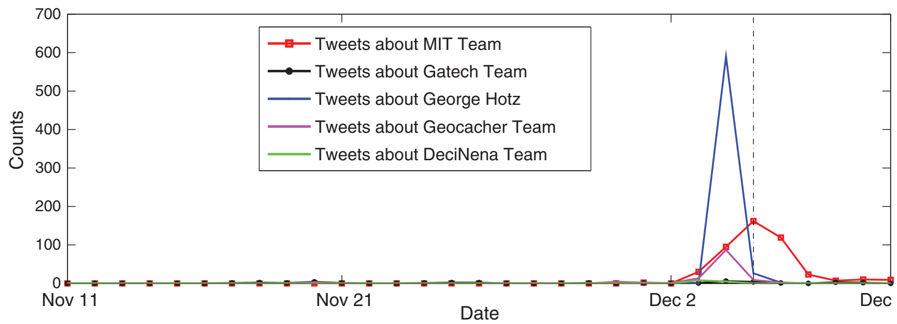


Fig. 4. Raw tweet counts for five teams from the announcement of the challenge to the announcement of the winner. The time series starts at the announcement of the challenge and ends at the announcement of the winner. The dotted line marks the time at which the balloons were launched. The MIT team launched its Web site and mechanism only 2 days before the balloon launch.

Because the DARPA challenge was not designed specifically as an experiment, the potential for comparison with the other teams is limited. To provide a qualitative comparison of diffusion between the MIT team and other teams, we analyzed data from the information network Twitter. We obtained ~100 million tweets for the time period from 10 November to 9 December. This data set covers an estimated 20 to 30% of all public tweets for that period (*31*). Initially, we filtered out all tweets except those containing the string "balloon" in a case-insensitive manner. We analyzed five teams in the top 10 final standings with a Twitter presence: MIT, GaTech, Hotz, Geocatcher, and Deci/Nena, representing different strategy categories. We then kept track of the number of tweets that included either of the following as tweet content about the team: team name, team website, hashtag for the team, short link for the team Web site, and team's affiliation name, including the abbreviation. The tweet counts are shown in Fig. 4.

The GaTech team adopted an altruism-based incentive method, by offering to donate all proceeds to the American Red Cross. The limited number of tweets responding to their strategy suggests that relying purely on altruistic propagation is not sufficient to amass large social mobilization. Because of their early start, mass media coverage, and search engine optimization, they ended up locating nine balloons with 1400 active members, and ranked second in the final list (*19*).

Another class of strategies is those that capitalize on an existing community of interest to which a team had direct access. We refer to this as the community-based strategy. George Hotz is a Twitter celebrity with more than 35,000 followers, and his strategy was to use his fame on Twitter to solicit help. He successfully created a burst in Twitter on the day he announced his participation in the competition, and ended up finding eight balloons (four from his Twitter network, and four through trades with other teams) (*17*). Similarly, Geocacher's strategy was based on the existing community of geocaching, a sport based on using navigational techniques to hide and seek objects. It also created a burst by announcing its participation to the geocacher community and located seven correct balloons. DeciNena aimed at assembling a balloon-hunting team by posting their participation on every related blog on the Internet to gain attention, but they failed to achieve a wide-range response. DeciNena found seven balloons at the end of the competition.

Although Hotz and Geocacher were able to create a sudden response peak by efficiently propagating the news to an existing audience, this response was very short-lived. On the other hand, our strategy was able to sustain social response for a longer period, stretching up until the end of the competition. This happened despite not having access to a large community of followers. Instead, the MIT team started with only four people; and after a couple of days, twitter response achieved a number comparable with that of Hotz, who started with 35,000 existing followers. Another interesting observation is that after the competition, when mass media came to report the winning story of the MIT team the tweet count actually decreased instead of increasing. This suggests that the incentives provided by the MIT strategy played a dominant role in generating Twitter response, rather than the "MIT brand" and mass media effect (SOM text).

The recursive incentive mechanism has a number of desirable properties. First, the recursive incentive mechanism is never in deficit—it never exceeds its budget (SOM text). After being recruited by a friend, an individual has no incentive to create his own root node by visiting the Balloon Challenge Web page directly (without using the link provided by the recruiter). This follows from the fact that payment to the person finding the balloon does not depend on the length of the chain of recruiters leading to him.

However, the mechanism is not resistant to false name attacks, which were originally identified in the context of Web-based auctions (*32*). In this attack, which has been shown to plague powerful economic mechanisms (*32*), an individual creates multiple false identities in order to gain an unfair advantage. Having said that, our data does not reveal any successful incidents of false-name attacks. This may be due to the fact that the mechanism did not operate for long enough for people to identify this potential, and that actual payment requires social security numbers. In practice, other measures could be put in place to minimize or detect this kind of attack (*33*).

The mechanism's success can be attributed to its ability to provide incentives for individuals to both reports on found balloon locations while simultaneously participating in the dissemination of information about the cause. When an individual finds a balloon, the individual can either report the balloon to us, to other teams, or attempt to find the other nine balloons and win the DARPA prize directly. In practice, it is unlikely for an unprepared individual to find other balloons (and if they replicated our mechanism, their delayed start would always leave them behind). Proofs are in the SOM.

Our mechanism simultaneously provides incentives for participation and for recruiting more individuals to the cause. This mechanism can be applied in very different contexts, such as social mobilization to fight world hunger, in games of cooperation and prediction, and for marketing campaigns.

References and Notes

1. J. Howe, *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business* (Three Rivers Press, New York, 2009).
2. E. Hand, *Nature* **466**, 685 (2010).
3. L. von Ahn, *Computer* **39**, 92 (2006).
4. S. Cooper *et al.*, *Nature* **466**, 756 (2010).
5. J. Pontin, *New York Times*, Artificial intelligence, with help from the humans (25 March 2007); available at www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html.
6. K. J. Arrow *et al.*, *Science* **320**, 877 (2008).
7. B. A. Huberman, D. M. Romero, F. Wu, *J. Inf. Sci.* **35**, 758 (2009).
8. G. Krakow, "Ham radio operators to the rescue after Katrina: Amateur radio networks help victims of the hurricane," www.msnbc.msn.com/id/9228945 (2005).
9. S. Milgram, *Psychol. Today* **1**, 6067 (1967).
10. J. M. Kleinberg, *Nature* **406**, 845 (2000).
11. D. J. Watts, P. S. Dodds, M. E. J. Newman, *Science* **296**, 1302 (2002).
12. L. A. Adamic, E. Adar, *Soc. Networks* **27**, 187 (2005).
13. M. Rosvall, P. Minnhagen, K. Sneppen, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **71**, 66111 (2005).
14. P. S. Dodds, R. Muhamad, D. J. Watts, *Science* **301**, 827 (2003).
15. W. Mason, D. J. Watts, *Proceedings of the ACM SIGKDD Workshop on Human Computation* (ACM, New York, 2009), pp. 77–85.
16. Defense Advanced Research Projects Agency, DARPA Network Challenge (accessed May 2010); available at http://networkchallenge.darpa.mil.
17. Defense Advanced Research Projects Agency, DARPA Network Challenge Project Report (February 16, 2010); available at www.hsdl.org/?view&did=17522.
18. Defense Advanced Research Projects Agency, "MIT Red Balloon Team wins DARPA Network Challenge" (press release, 5 December 2009).
19. J. C. Tang *et al.*, *Commun. ACM* **54**, 78 (2010).
20. D. J. Watts, J. Peretti, *Harv. Bus. Rev.* **May**, 22 (2007).
21. J. Kleinberg, P. Raghavan, in *Proccedings of 46th Annual IEEE Symposium on FOCS* (IEEE, Los Alamitos, CA, 2005), pp. 132–141.
22. E. Biyalogorsky, E. Gerstner, B. Libai, *Mark. Sci.* **20**, 82 (2001).
23. J. Iribarren, E. Moro, *Phys. Rev. Lett.* **103**, 038702 (2009).
24. D. J. Watts, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5766 (2002).
25. D. Liben-Nowell, J. Kleinberg, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4633 (2008).
26. B. Golub, M. O. Jackson, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10833 (2010).
27. Our data does not include the time stamp of sending out invitations, but we were able to measure the intervals between actual sign-up events between a parent and its children in the trees.
28. A. L. Barabási, *Nature* **435**, 207 (2005).
29. R. D. Malmgren, D. B. Stouffer, A. S. Campanharo, L. A. Amaral, *Science* **325**, 1696 (2009).
30. D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11623 (2005).
31. J. Yang, J. Leskovec, *Proceedings of the 4th ACM International Conference on Web Search and Data Mining* (ACM, Kowloon, Hong Kong, 2011).
32. M. Yokoo, Y. Sakurai, S. Matsubara, *Games Econ. Behav.* **46**, 174 (2004).
33. D. Whitworth, Man fined over fake eBay auctions. *BBC Newsbeat* (5 July 2010); available at www.bbc.co.uk/newsbeat/10508913.

Acknowledgments: This research was partially sponsored by the Army Research Laboratory under cooperative agreement W911NF-09-2-0053, by the Air Force Office of Scientific Research under award FA9550-10-1-0122, by the U.S. Air Force under contract FA8721-05-C-0002, by the National Science Foundation under grant 0905645, by the Masdar Institute of Science and Technology under grant 400075, the Branco Weiss Science in Society Program, and by generous gifts from P&G, Best Buy, and Google Views. Conclusions in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the sponsors. After providing the data underlying Fig. 4, Twitter requested that Stanford University withdraw this data from public access. Please contact A.P. for information regarding use of the data on the MIT server.

Supporting Online Material
www.sciencemag.org/cgi/content/full/334/6055/509/DC1
Materials and Methods
SOM Text
Figs. S1 to S5

21 March 2011; accepted 15 September 2011
10.1126/science.1205869

The Complex Folding Network of Single Calmodulin Molecules

Johannes Stigler,[1] Fabian Ziegler,[1] Anja Gieseke,[1] J. Christof M. Gebhardt,[1]* Matthias Rief[1,2]†

Direct observation of the detailed conformational fluctuations of a single protein molecule en route to its folded state has so far been realized only in silico. We have used single-molecule force spectroscopy to study the folding transitions of single calmodulin molecules. High-resolution optical tweezers assays in combination with hidden Markov analysis reveal a complex network of on- and off-pathway intermediates. Cooperative and anticooperative interactions across domain boundaries can be observed directly. The folding network involves four intermediates. Two off-pathway intermediates exhibit non-native interdomain interactions and compete with the ultrafast productive folding pathway.

The energy landscape view provides a conceptual framework for understanding protein folding (*1*, *2*). However, the diversity in size and structure of the proteome is far too large to provide a single generic mechanism for how proteins fold. Deciphering specific mechanisms

www.sciencemag.org/cgi/content/full/334/6055/509/DC1

Supporting Online Material for

Time-Critical Social Mobilization

Galen Pickard, Wei Pan, Iyad Rahwan, Manuel Cebrian, Riley Crane, Anmol Madan, Alex Pentland*

*To whom correspondence should be addressed. E-mail: pentland@mit.edu

Published 28 October 2011, *Science* **334**, 509 (2011)
DOI: 10.1126/science.1205869

This PDF file includes:

Materials and Methods

SOM Text

Figs. S1 to S5

Time Critical Social Mobilization: Supporting Online Material

Galen Pickard,[1,2,†] Wei Pan,[1†] Iyad Rahwan,[3,1,†] Manuel Cebrian,[1,†]
Riley Crane,[1] Anmol Madan,[1] Alex Pentland[1,‡]

[1]Massachusetts Institute of Technology, USA
[2]Massachusetts Institute of Technology Lincoln Laboratory, USA
[3]Masdar Institute of Science & Technology, UAE

[†]These authors contributed equally to the paper.
[‡]Corresponding author; E-mail: `pentland@mit.edu`

September 16, 2011

1 Materials and Methods

We implemented the MIT mechanism by building a simple web platform (`http://balloon.media.mit.edu`). Any user could sign up with an E-mail address to become our team member and report a balloon on the web platform. During the registration and balloon reporting process, we recorded the time-stamp and IP address for each user activity on the platform. After registration, the user received a unique link in an E-mail from us, and the user could refer others by asking their friends to sign up with the MIT team via this link. The link allowed our system to record the referral relationships between users. It was up to the users to decide how to share this unique link. Some users chose to e-mail the link to their friends, but many others shared the link in online social media tools such as Twitter and Facebook. We initialized the diffusion process by sending the link to our web platform to a few close friends of the team members and several online web blogs 2 days before the competition.

Our analysis in Figure 1 and Figure 2 is based on the sign-up time information and referral chains collected from our web platform. We translated all user IP addresses into their locations using a 3rd-party online database *smart-ip.net*. Based on this location information, we were able to perform the analysis in Figure 3.

All the raw data including time stamps collected by our platform during the DARPA Challenge as well as the translated IP addresses are available as MySQL databases from the authors.

The raw Twitter dataset was collected in a Stanford research project by Yang *et al.*(31), and this project is unrelated to the DARPA Challenge. The Stanford researchers monitored and recorded the public Twitter feed from June, 2009 to December, 2009, and the crawled Twitter dataset contains 20-30% of all public tweets during their study period. The dataset had been available to public until recently when Twitter requested Stanford to withdraw from public access[1]. We used this Twitter dataset to scan for related tweets with keywords described in the main text, and conducted the analysis in Figure 4.

Other teams mentioned in our main text used different strategies to diffuse their involvement and to recruit their members: The GaTech team created their own website `http://www.ispyaredballoon.com`, embedded links towards its website on Georgia Tech and Georgia Tech Research Institute websites, and performed search engine optimization to attract traffic. They also received a National Public Radio *"Here and Now"* broadcast interview four days before the challenge (17). George Hotz simply communicated his participation on his Twitter feeds to his 50,000 followers one day before the challenge. The Geocacher team sent E-mail alerts to all the members in the Geocacher community (roughly several hundred thousand members) one and two days prior to the balloon launch (17). They also set up a blog (`http://www.10ballonies.com`) to post their progress. DeciNena posted their information (`http://decinena.com`) in the comment section of every related DARPA Challenge blog post they could find online starting three days prior to the balloon launch[2].

2 Formal Definition for the MIT Strategy

The MIT team approach was based on the idea that achieving large-scale mobilization towards a task requires diffusion of information about the tasks through social networks, as well as incentives for individuals to act, *both* towards the task and towards the recruitment of other individuals.

We consider our approach to the DARPA Network Challenge to be an instance of a more general class of mechanisms for distributed task execution. We define a *diffusion-based task environment* which consists of the following: $N = \{\alpha_1, \dots, \alpha_n\}$ is a set of *agents*; $E \subseteq N \times N$ is a set of *edges* characterizing social relationships between agents; $\Psi = \{\psi_1, \dots, \psi_m\}$ is a set of *tasks*; $P : N \times \Psi \to [0, 1]$ returns the *success probability* of a given agent in executing a given task; $B \in \mathbb{R}$ be the *budget* that can be spent by the mechanism.

[1]See `http://snap.stanford.edu/data/twitter7.html`.

[2]We estimated their start time according to the information on their Facebook post: `http://www.facebook.com/pages/Team-DeciNena/192350616581`.

In a diffusion-based task environment, unlike in traditional task allocation mechanisms (e.g. based on auctions), agents are *not* aware of the tasks a priori. Instead, they *become* aware of tasks as a result of either 1) being directly informed by the mechanism through advertising; or 2) being informed through recruitment by an acquaintance agent (20).

Another characteristic of diffusion-based task environments is that, when a task is completed, the mechanism is able to identify not only the agent who executed it, but also the information pathway that led to that agent learning about the task. The pathway leading to the successful completion of task ψ_i is captured by the sequence $\mathcal{S}(\psi_i) = \langle a_1, \ldots, a_r \rangle$ of unique agents, where a_r is the agent who completed the task, a_r was informed of the task by a_{r-1} and so on up to agent a_1 who was initially informed of the task by the mechanism. By slightly overloading notation, let $|\mathcal{S}(\psi_i)|$ denote the length of the sequence (i.e. the number of agents in the chain), and let $\alpha_j \in \mathcal{S}(\psi_i)$ denote that agent α_j appears in sequence $\mathcal{S}(\psi_i)$.

We can now define a class of mechanisms that operate in the above settings. A *diffusion-based task execution mechanism* specifies the following: $I \subseteq N$ is a set of *initial nodes* to target (e.g. via advertising); ρ_i is the *payment* made to agent α_i; such that the following constraint is satisfied: $c|I| + \sum_{\alpha_i \in N} \rho_i \leq B$.

In words, the mechanism makes two decisions. First, it decides which nodes to target initially via advertising. Second, it decides on the payment (if any) to be made each agent. The mechanism must do this within its budget B.

In the DARPA Network Challenge, each ψ_i represents finding a balloon, and $v(\psi_i) = 4,000$ for all $\psi_i \in \Psi$. Moreover, we assume that the ten tasks are all identical (namely finding a balloon), and all task are indistinguishable, $\forall \alpha_i \in N, \forall \psi_k, \psi_l \in \Psi$ we have $P(\alpha_i, \psi_k) = P(\alpha_i, \psi_l)$. That is, the success probability of a particular agent is the same for all balloons.

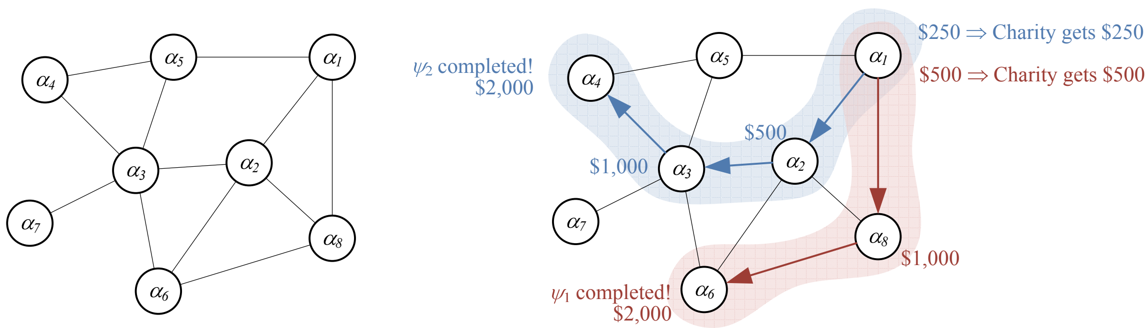
We are now ready to define our mechanism, referred to as a *recursive incentive mechanism*. Given I initial targets, and assuming $v(\psi_i) = B/|\Psi|$, divide the budget B such that each task $\psi_i \in \Psi$ has budget $B_i = B/|\Psi|$. If agent $j \in N$ appears in position k in sequence $\mathcal{S}(\psi_i)$, then j receives the following payment:

$$\frac{v(\psi_i)}{2^{(|\mathcal{S}(\psi_i)|-k+1)}} \tag{1}$$

Hence, the total payment received by agent j is the sum of payments for all sequences in which j appears:

$$\rho_j = \sum_{\psi_i | j \in \mathcal{S}(\psi_i)} \frac{v(\psi_i)}{2^{(|\mathcal{S}(\psi_i)|-k+1)}} \tag{2}$$

The surplus is therefore: $S = B - \sum_{\alpha_j \in N} \rho_j$. Figure S1 illustrates how this mechanism works.



(a) Example social network.

(b) Recruitment tree with two paths (shown in thick lines) initiated by α_1 led to finding balloons.

Figure S1: Recursive incentive mechanism: (a) Suppose that in this network, agent α_1 recruits all of his neighbors, namely α_2, α_5 and α_8. Suppose that α_8 recruits α_6, who finds balloon ψ_1. (b) We have a winning sequence $\mathcal{S}(\psi_1) = \langle \alpha_1, \alpha_8, \alpha_6 \rangle$ with $|\mathcal{S}(\psi_1)| = 3$. The finder receives $\rho_8 = \frac{4{,}000}{2^{(3-3+1)}} = 2,000$. Since α_8 recruited α_6, then $\rho_8 = \frac{4{,}000}{2^{(3-2+1)}} = 1,000$. From this sequence, α_1 receives $\frac{4{,}000}{2^{(3-1+1)}} = 500$. Likewise, looking at the left recruitment path, we have a winning sequence $\mathcal{S}(\psi_2) = \langle \alpha_1, \alpha_2, \alpha_3, \alpha_4 \rangle$ with $|\mathcal{S}(\psi_2)| = 4$. The finder receives $\rho_4 = \frac{4{,}000}{2^{(4-4+1)}} = 2,000$. As above, we have $\rho_3 = \frac{4{,}000}{2^{(4-3+1)}} = 1,000$ and $\rho_2 = \frac{4{,}000}{2^{(4-2+1)}} = 500$. From this sequence, α_1 receives $\frac{4{,}000}{2^{(4-1+1)}} = 250$. Adding up its payments from the two sequences it initiated, α_1 receives a total payment of $\rho_1 = 750$. Assuming there are only two tasks, the surplus in this case is $S = (4,000 - 3,500) + (4,000 - 3,750) = 750$.

3 Formal Proofs

3.1 Mechanism Always within Budget

Proposition 1. *The recursive incentive mechanism is never in deficit (i.e. never exceeds its budget).*

Proof. Recall that each sub-task ψ_i is allocated an equal share of $B_i = B/|\Psi|$ budget. Hence, it suffices to show that the payment for any arbitrary task ψ_i is bounded by B_i. Let $\mathcal{S}(\psi_i) = \langle a_1, \ldots, a_r \rangle$ be the (finite) sequence leading to the successful completion of ψ_i. We need to show that the total payment made to all agents in sequence $\mathcal{S}(\psi_i)$ within budget, that is, we need to show that:

$$\sum_{k=1}^{r} \frac{B_i}{2^{(r-k+1)}} \leq B_i \quad \textit{or equivalently we need to show that} \quad \sum_{k=1}^{r} \frac{1}{2^{(r-k+1)}} \leq 1$$

We can easily see that: $\sum_{k=1}^{r} \frac{1}{2^{(r-k+1)}} = \sum_{k=1}^{r} (\frac{1}{2})^{(r-k+1)} = \sum_{k=1}^{r} \frac{1}{2} \times (\frac{1}{2})^{(r-k)}$

Defining $i = r - k$, we can rewrite:

$$\begin{aligned} \sum_{k=1}^{r} 0.5(\tfrac{1}{2})^{(r-k)} &= 0.5(\tfrac{1}{2})^{(r-1)} + 0.5(\tfrac{1}{2})^{(r-2)} + \ldots 0.5(\tfrac{1}{2})^{(r-r)} \\ &= 0.5(\tfrac{1}{2})^{(r-1)} + 0.5(\tfrac{1}{2})^{(r-2)} + \ldots 0.5(\tfrac{1}{2})^{0} \\ &= \sum_{i=0}^{r-1} 0.5(\tfrac{1}{2})^{i} \end{aligned}$$

This is a finite geometric series, with a well-known closed form:

$$\sum_{i=0}^{r-1} 0.5(\tfrac{1}{2})^{i} = 0.5\frac{1-(\frac{1}{2})^{(r-1)+1}}{1-\frac{1}{2}} = 1 - (\tfrac{1}{2})^{r} = \frac{2^r - 1}{2^r} \leq 1 \qquad (\text{for} \quad r \geq 1)$$

□

We continue to discuss our theoretical analysis on the mechanism. When agent α_i becomes aware of a set of tasks $\{\psi_1, ..., \psi_m\}$, it needs to select a (possibly empty) set of neighbors $T(\alpha_i) \subseteq \{\alpha_j \in N \ : \ (\alpha_i, \alpha_j) \in E\}$ to recruit (i.e. to inform them about ψ), assuming the team will win the challenge. The diffusion of information about the task relies crucially on such recruitment choices among agents. One can perform this incentive analysis under two different assumptions:

3.2 Incentives With Uniform Independent Success Probability Among All Population

We now discuss the properties our mechanism under the assumption that the probability of each person finding a balloon is an independent (and very small) constant, $\forall i, k, P(\alpha_i, \psi_k) = \epsilon$, such that $n.\epsilon \leq 1$, i.e. the sum of these probabilities over the *entire* population (including those not recruited) is bounded by 1. In this case, it is trivial to show that recruiting all of one's peers is the best strategy. Without recruiting, one achieves an expected reward of $\sum_i \epsilon \frac{v(\psi_i)}{2}$. With recruiting, on the other hand, one's expected rewards is $\sum_i \left(\epsilon \frac{v(\psi_i)}{2} + \sum_j \epsilon x_j \frac{v(\psi_i)}{2^j}\right)$, where x_j is the number of individuals at depth j of the recruiter's tree. Clearly, this expected reward increases monotonically in the number of directly recruited nodes.

3.3 Incentives With Uniform Success Probability Among Recruited Individuals

We can also analyze incentives under the assumption that the probability of an individual finding a balloon is uniformly distributed across the *recruited* individuals. Formally we have:

Definition 1 (Uniformity)**.** *Each recruited individual is equally likely to accomplish task ψ, and the probability of each individual accomplishing the task is $\frac{1}{n}$, where n is the number of all recruited individuals.*

Intuitively, it means that a fixed-size group of recruited individuals is guaranteed to find the balloon eventually, even if no other individuals are recruited. This assumption is realistic if the set of recruited individuals is sufficiently large (e.g. thousands). We will continue to show that, under fairly broad assumptions on the structure of the society, it is also in the best interest of each individual to recruit all their friends. In particular, here we show that if no individual controls $n/3$ of the population (i.e. is able to prevent them from learning about the task), then the strategy profile in which all individuals recruit all their friends is a Nash equilibrium. In the following results, we consider the situation that there is only one task ψ that is being diffused in the social network with a total budget of 1.

3.3.1 All-or-None Recruitment on Fixed-Forest Social Networks

We consider the case in which the social network takes the form of a forest of rooted trees, and the roots of these trees form the set of initially-recruited nodes I.

Given this forest F, which contains a total of n nodes, each node chooses whether or not to recruit all of its children. This induces a "recruited subforest" F' of size n', consisting of all nodes which can trace a path of recruitment to a root node of F.

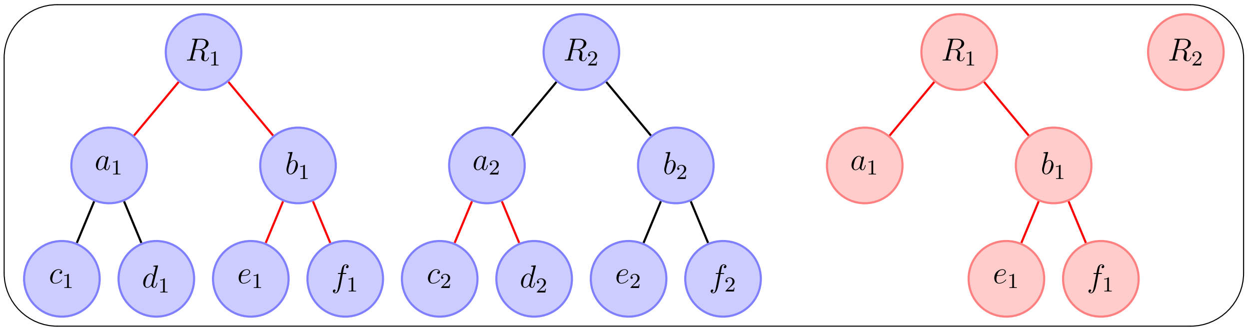


Figure S2: Nodes R_1, b_1, and a_2 choose to recruit; the rest not. The recruited subforest F' is shown in red. Note that a_2's choice to recruit is rendered moot by R_2's choice not to recruit.

For each node in the recruited subforest, this results in an expected payment based solely on the shape of its descendant recruited subtree. For each node, we can characterize this shape with an ordered tuple $X = \langle x_1, x_2, x_3, \ldots 0\rangle$, representing the number of children, grandchildren, great-grandchildren, etc. (i.e. in the example, R_1's tuple would be $\langle 2, 2, 0\rangle$, and the tuple of any leaf node is $\langle 0\rangle$). Given such a tuple, under the uniformity assumption, the expected payment to a node is

$$U(X) = \frac{1 + \sum_i \frac{x_i}{2^i}}{n'},$$

where n' is the number of nodes in the recruited subforest.

Given the set of choices (recruit all children or recruit no children) made by each node, this function $U(X)$ is a payout function which defines a normal-form game played by all non-leaf nodes in the original forest.

3.3.2 Game Definition

We demonstrate the definition of the game by example, recreating the "prisoner's dilemma" using a 5-node forest under the uniformity assumption.

Consider the forest F shown in Figure S3. There are two players, R_1 and R_2, each of which has the option to recruit a single child or not. If neither recruits, both receive an expected payment of $\frac{1}{3}$. If one recruits but the other does not, the recruiter has an expected payment of $\frac{1+\frac{1}{2}}{4} = \frac{3}{8}$, while the other has an expected payment of $\frac{1}{4}$. If both recruit, both have an expected payment of $\frac{1+\frac{1}{2}}{5} = \frac{3}{10}$. This gives a payment matrix approximated by:

	N	Y
N	.33, .33	.25, .37
Y	.37, .25	.3, .3

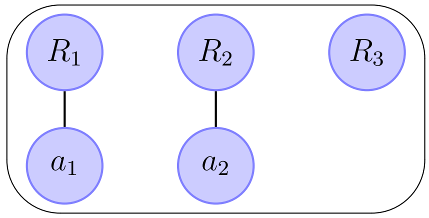


Figure S3: The game played by R_1 and R_2 is equivalent to the "prisoner's dilemma."

Clearly, choosing to recruit is a strictly dominant strategy for each player, so the only Nash equilibrium that both players recruit – even though this is Pareto inefficient.

3.3.3 Nash Equilibrium of Larger Forests

Consider a game in which all actors have two options: "recruit all" or "recruit none." For any given agent a, let all other agents choose "recruit all," and consider a's optimal strategy. If choosing "recruit all" is optimal for a, then no agent can benefit by deviating from a strategy of "recruit all," if all other agents choose "recruit all." This, by definition, makes the uniform choice to "recruit all" a Nash equilibrium.

Theorem 1. *A node a will prefer recruitment to non-recruitment predicated on all other nodes choosing to recruit if and only if sufficiently many nodes in the forest F are not descendants of a under the uniformity assumption.*

Proof. For a node a in a forest F of size n, let the tuple $X = \{x_1, x_2, x_3, \ldots\}$ be defined as the number of children, grand-children, great-grand-children, etc. of node a. If F is finite, each x_i is finite and there exists some j such that $x_i = 0$ for all $i > j$. Let k be the number of nodes in F that are not descendants of a, noting that $k = n - \sum_i x_i$. Since we assume all nodes other than a choose to recruit, the expected payment received by a if a chooses to recruit is $\frac{1}{k}$. If a does choose to recruit, then a will receive expected payment $\frac{1+\sum_i \frac{x_i}{2^i}}{n} = \frac{1+\sum_i \frac{x_i}{2^i}}{k+\sum_i x_i}$. a will find it preferable to recruit if and only if $\frac{1+\sum_i \frac{x_i}{2^i}}{k+\sum_i x_i} > \frac{1}{k}$, or, equivalently, when $k > \frac{\sum_i x_i}{\sum_i \frac{x_i}{2^i}}$.

□

Corollary 1. *In any forest F of size n for which no tree contains more than $\frac{n}{3}$ nodes, all nodes choosing to recruit is a Nash equilibrium under the uniformity assumption.*

Proof. Consider forest F with n nodes, and a node a which has m descendants, taking a shape described by a tuple $X = \{x_1, x_2, x_3, \ldots\}$. The expected payment for a for not recruiting is $\frac{1}{n-m}$, and the expected payment for a for recruiting is $\frac{1+\sum_i \frac{x_i}{2^i}}{n}$. Therefore, we have that a will choose to recruit predicated on all other nodes recruiting if and only if $n - m > \frac{m}{\sum_i \frac{x_i}{2^i}}$. We note that the definition of X yields that no non-zero value can follow a zero value (i.e. one must have grand-children in order to have great-grand-children). It follows that, if we fix m, the setting of X which maximizes $\frac{m}{\sum_i \frac{x_i}{2^i}}$ is $X = \{\overbrace{1, 1, \ldots, 1}^{m}, 0, 0, \ldots\}$, which gives $\frac{1}{2} \leq \sum_i \frac{x_i}{2^i} < 1$ for $m > 0$. Thus, a will choose to recruit if (but not only if) $n - m > 2m$. This condition holds for all nodes if and only if no tree in F contains more than $\frac{n}{3}$ nodes. In this case, all nodes will choose to recruit predicated on all other nodes recruiting, so all nodes choosing to recruit is a Nash equilibrium.

□

3.3.4 Selective Recruitment on Fixed-Forest Networks

We now consider the same social graph structure, but allow a node to selectively recruit any subset of its children.[3]

Definition 2 (Weight)**.** *We define the weight of a node a, W_a, as the sum of the rewards that would be received by a in the event that each of its descendants were to complete the task. We note the following properties of W_a*

- *If a is a leaf, then $W_a = 1$.*
- *If a has children $c_1, c_2, \ldots$ with weights $W_{c_1}, W_{c_2}, \ldots$, then $W_a = 1 + \frac{1}{2}\sum_i W_{c_i}$.*
- *If node a has descendants described by shape $X = \langle x_1, x_2, \ldots, 0\rangle$, then $W_a = 1 + \sum_i \frac{x_i}{2^i}$.*
- *In a forest with n nodes, the expected payment to node a is $U(a) = \frac{W_a}{n}$.*

Lemma 1. *Assuming all other nodes recruit all their children, a node a with children $C = \{c_1, \ldots, c_m\}$ recruits a child c_x regardless of the shape of F_{c_i} of the other children $c_i \neq c_x$, if and only if the weight of c_x is large relative to the number of its descendants: $\frac{\frac{1}{2}(m+1)}{k+m-1} \leq \frac{\frac{1}{2}W_{c_x}}{|c_x|}$*

Proof. Let $k > 2$ be the number of nodes in the forest that are not descendants of a. The marginal benefit of recruiting a child c_x of node a depends on the other children a recruits. Assuming a recruits a subset of its children $S \subset (C \setminus \{c_x\})$ and does not recruit any other children $C \setminus (S \cup \{c_x\})$, child c_x is recruited if and only if

$$\frac{1 + \frac{1}{2}\sum_{c_i \in S} W_{c_i}}{k + (\sum_{c_i \in S} |c_i|)} \leq \frac{1 + \frac{1}{2}\sum_{c_i \in S} W_{c_i} + \frac{1}{2}W_{c_x}}{k + (\sum_{c_i \in S} |c_i|) + |c_x|}$$

[3] We are grateful to Victor Naroditskiy for comments that helped refine the results in this section.

This inequality is equivalent to

$$\frac{1+\frac{1}{2}\sum_{c_i\in S} W_{c_i}}{k+(\sum_{c_i\in S}|c_i|)} \leq \frac{\frac{1}{2}W_{c_x}}{|c_x|}$$

We are looking for a condition that guarantees c_x is recruited for any subset of S and shapes of trees F_{c_i} rooted at children other than c_x. Thus, we want the left-hand side to be as high as possible. For $k > 2$, the left-hand side is maximized when $S = C \setminus \{c_x\}$ and each child $c_i \in S$ has no descendants. This worst-case value is

$$\frac{1+\frac{1}{2}(m-1)}{k+(m-1)} = \frac{\frac{1}{2}(m+1)}{k+m-1}$$

□

Corollary 2. *Node a recruits all of its children if Lemma 1 holds for each child.*

Theorem 2. *Assuming all other nodes recruit all their children, a node a recruits all of its children regardless of the shape of F_a if sufficiently many nodes are not a's descendants: $k > |a|^2$*

Proof. The condition $k > |a|^2$ is equivalent to $\frac{1}{|a|} > \frac{|a|}{k}$. Since $|c_x| \leq |a|$, we have $\frac{1}{|c_x|} > \frac{|a|}{k}$. For $a \geq m+1$ (for the case $a = m$, all children of a have no descendants, and it is easy to see that they are all recruited if $k > 2$), we obtain $\frac{1}{|c_x|} > \frac{m+1}{k}$. Trivially, $W_{c_x} > 1$, and thus $\frac{W_{c_x}}{|c_x|} > \frac{m+1}{k} > \frac{m+1}{k+m-1}$ and by Lemma 1 c_x should be recruited. □

3.3.5 Recruitment on Graphs

We consider now the case in which the social graph is not a forest, but is instead a general graph. In this case, the mechanism of recruitment itself plays a non-trivial role, since it is possible for a node to be recruited by two different potential parents, and must choose between them. There is significant literature on diffusion processes on graphs, and wide varieties of such processes are seen in practice. We will not investigate the properties of specific diffusion mechanisms, but instead we will define a property of a diffusion mechanism that guarantees that recruitment is Nash.

Definition 3 (Monotonic Diffusion)**.** *Consider a diffusion process on a social graph, and a set of seed nodes $R_1, R_2, \ldots, R_n$. Let $|R_1|, |R_2|, \ldots, |R_n|$ be the number of nodes whose recruitment leads back to $R_1, R_2, \ldots, R_n$, respectively. We call the diffusion process monotonic if removing a seed node R_x causes the sizes of $|R_1|, |R_2|, \ldots, |R_{x-1}|, |R_{x+1}|, \ldots, |R_n|$ to either increase or stay constant (i.e. if R_x does not participate, this does not cause another seed node to recruit fewer children).*

Monotonicity holds for most "well-behaved" diffusion processes, but is notably violated by various "complex contagion" processes in which, for example, a node adopts after receiving two signals.

Theorem 3. *Under the monotonic diffusion assumption, if each node can recruit less than $\sqrt{k}$ nodes, where k is the number of all recruited nodes that are not descendant of this node when all nodes recruit, then all nodes recruiting is a Nash equilibrium.*

Proof. Consider a node a, which can choose whether or not to recruit, and suppose all other nodes recruit. Suppose it were the case that if a were to not choose recruitment, then all nodes that would have been recruited by a would end up un-recruited, instead. In this case, the graph reduces to the same fixed forest we analyzed previously. Suppose instead that some of these nodes end up recruited by a different node. In this case, not recruiting is strictly less desirable, since the size of the network grows without any increase in potential payout. Hence, it follows Theorem 2 that recruiting all children is more desirable in either case under the assumption that a cannot recruit more than $\sqrt{k}$. If diffusion is monotonic, the two cases considered are collectively exhaustive, so recruiting all children is always the more desirable option.

□

3.4 Summary

The two assumptions above differ in their treatment of how the addition of a new member to the network affects the probability that each other member succeeds in finding a balloon. The first assumption is that new members have no effect on existing members, perhaps because they are searching mutually exclusive areas, and if the new member were to find a balloon, that implies that no-one would have found that balloon in his absence. The second assumption is that each member's probability decreases from $\frac{1}{n}$ to $\frac{1}{n+1}$, perhaps because all members share the same search space. Unless "network effects" are present (e.g. working with a new member makes us both more likely to succeed than working alone), these two assumptions represent best- and worst-case assumptions. There are certainly intermediate assumptions that could be made, and the strategies that motivate diffusion at both extremes will do so for these intermediate assumptions as well.

4 Comparison with Query Incentive Networks

We consider how our scheme relates to Kleinberg and Raghavan pioneering work on "Query Incentive Networks" (QINs) (21). A QIN is a network of agents which one agent seeks information which is known to some subset of other agents. This agent broadcasts to all of his neighbors that he is offering a reward of value r^* in exchange for the information. Assuming that none of his neighbors have the information themselves, each can broadcast to its own neighbors that it is offering a lesser reward (e.g. r^*-1) for the same information. If it receives the information from any of its neighbors, it pays out $r^* - 1$ and receives r^* from the root node, making a profit of 1 by acting as a conduit for the information. Kleinberg et al. analyze the properties of these networks, and show that they are an efficient method for retrieving uncommon information in a network of agents. QINs are very similar in spirit to our recursive incentives, and we show how recursive incentives can be created through a modification to the QIN strategy. In addition, we argue that recursive incentives effectively solve two issues that would make practical implementation of a QIN difficult in a time-critical situation.

When comparing recursive incentives and QINs, two differences are paramount. First, in a recursive incentive network, upon receipt of the information, the root node directly pays each node on the path to the agent who supplied the information, while in a QIN, the root node makes a single payment to one of its neighbors, who then makes a smaller payment to one of its neighbors, and so on. In the construction of the recursive incentive network, then, is the implicit assumption that any node can communicate (and transfer payment) to any other, but that the propagation of queries can only follow a limited number of links (i.e. the network structure). This assumption is a natural model of real social networks (especially on-line social networks like Facebook or Twitter) in which it is possible to explicitly communicate with any user by name, but the default "broadcast" mechanism only reaches a limited, socially-defined subset of users. A second difference is that in a QIN, intermediate nodes on the chain to the information supplier receive a fixed reward, and the node at the end of the chain receives an amount that decreases as its distance from the root increases. By contrast, in a recursive incentive network, the node that supplies the information receives a fixed reward, while intermediate nodes receive a variable reward depending on their distance from the information supplier.

4.1 Making Recursive Incentives out of Query Incentive Networks

We borrow terminology from Kleinberg et al. to show how to transform the QIN framework into a recursive incentive. We say that the root node offers a "contract" of value r^* for a piece of information, effectively promising "I will pay r^* to the first node to provide me with this information." Other nodes that do not themselves value the information offer "subcontracts" of value less than r^*, hoping to receive reward by acting as an intermediary. Consider if the root node adds the following to the contract: "In addition to paying r^* for the information, if you obtain this information though subcontracts and your subcontracts are identical to this contract, I will reimburse you for half of the costs you pay." This results in a payment scheme that is identical to a recursive incentive which pays the final node r^*, as we will show.

The base case is the situation in which a neighbor (n_1) of the root has the information. The root node pays r^* to n_1. If n_1 does not have the information, but offers the same contract to its neighbor (n_2), who does have the information, n_1 pays r^* to n_2. The root then pays r^* to n_1 for fulfilling the original contract, and an additional $\frac{r^*}{2}$ to reimburse half of the costs paid by n_1. Consider the case in which n_2 does not have the information, but offers the same contract to n_3, who does. Working from the leaf node up to the root:

- n_2 pays r^* to n_3 for fulfilling the contract. n_3 makes no payments, so receives a net of r^*.
- n_1 pays r^* to n_2 for fulfilling the contract, and also pays $\frac{r^*}{2}$ to n_2, since n_3 paid out r^* in expenses. n_2 received $\frac{3r^*}{2}$ and paid out r^*, for a net profit of $\frac{r^*}{2}$.
- The root pays r^* to n_1 for fulfilling the contract, and also pays $\frac{3r^*}{4}$ to n_1, since n_1 paid out a total of $\frac{3r^*}{2}$. n_2 received $\frac{7r^*}{4}$ and paid out $\frac{3r^*}{2}$, for a net profit of $\frac{r^*}{4}$.

Having now explicitly demonstrated how small chains produce the same payouts as recursive incentives, let us now consider the case where the chain is of length k:

- n_{k-1} pays r^* to n_k for fulfilling the contract. n_k makes no payments, so receives a net of r^*.
- For i between 1 and $k-1$, inclusive, define p_1 as the total amount paid out by n_i. Note that the net profit of node n_i is $p_{i+1} - p_i$
- To fulfill all aspects of the contract, n_i pays $\frac{p_{i+1}}{2} + r^*$ to n_{i+1}, so $p_i = \frac{p_{i+1}}{2} + r^*$.
- Letting $p_{k-1} = r^*$, $p_{k-i} = \frac{r^* 2^i - 1}{2^{i-1}}$ is the unique solution to this recursive definition of p_i, which leads to node n_i making net profit $\frac{r^*}{2^{k-i}}$.

4.2 Practical Considerations

Having shown that recursive incentives can be presented as a modification to the concept of QINs, we argue that the recursive incentive formulation has significant benefits when implemented in practice. Consider a traditional QIN operating over a real-world online social network like Facebook or Twitter. When a node sees a contract and wants to offer a subcontract, it has two practical difficulties. First, it cannot allow any of its subcontractors to know about its original (and more valuable) contract, since any would-be subcontractor who knows about the original would surely prefer to fulfill it instead. Also, it must distribute the nature of the request and establish payment infrastructure (including building the trust in the subcontractors that payment will happen). Since it cannot use the original contract as a reference, it is effectively starting from scratch, which significantly increases the cost of initiating the subcontract. By contrast, propagation in a recursive incentive network involves simply passing a link to the original "contract," allowing the root node to bear all the burden of fully explaining the request and establishing payment infrastructure.

A second real-world issue faced by a QIN (especially in a time-critical situation) occurs when communication between nodes incurs a time delay or is lossy. In a QIN, the information in its entirety must pass through a series of intermediaries. If any of these corrupts the information or significantly delays propagation (e.g. goes offline), the information could potentially not reach the root node in a timely and correct fashion. By contrast, a recursive incentive network allows the node that has the information to pass it directly to the root node, bypassing all intermediaries (while still allowing them to be paid by the root). Again, this assumes "broadcast" communication is limited by social connection, but that point-to-point communication is possible between any two nodes in the network – but this assumption underlies the core dynamics of most online social networks.

5 Quantifying the "MIT" Effect

In the paper, we argued that the "MIT brand" did not play a major role in the success of the MIT team. In particular, we showed that the burst of tweets about the MIT team was more sustained compared to other strategies, including those based on celebrity following, which experienced very short-lived bursts. We attributed this qualitative difference to the MIT incentive mechanism.

To further support this claim, bearing in mind the limited data available, we compared the MIT red balloon team's tweet count with another MIT-related event, namely launching a hybrid electrical bicycle[4]. This event took place in the same month, received significant mass media attention[5], and was the only MIT news exceeding 50 tweets in December in our Twitter dataset (31). While it is difficult to conduct a systematic comparison, Figure S4 suggests that this event also sustained a short-lived burst even with major media coverage, while the MIT team achieved sustaining burst with our mechanism.

[4]See `http://senseable.mit.edu/copenhagenwheel/`.

[5]E.g. see `http://www.nytimes.com/2009/12/15/science/earth/15bike.html`

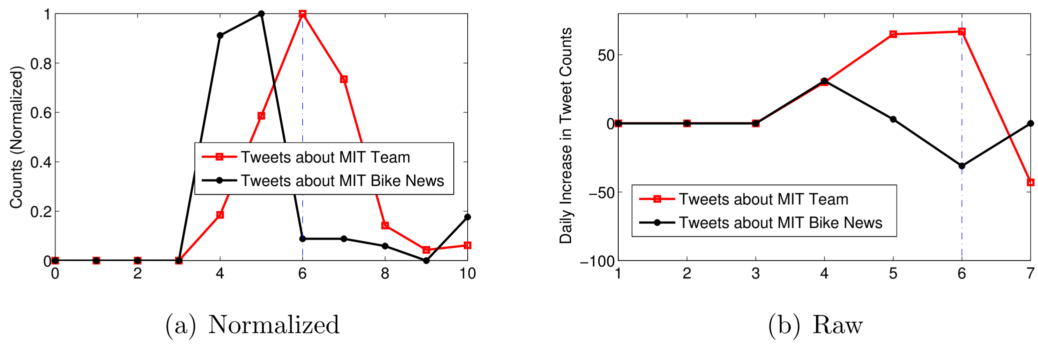
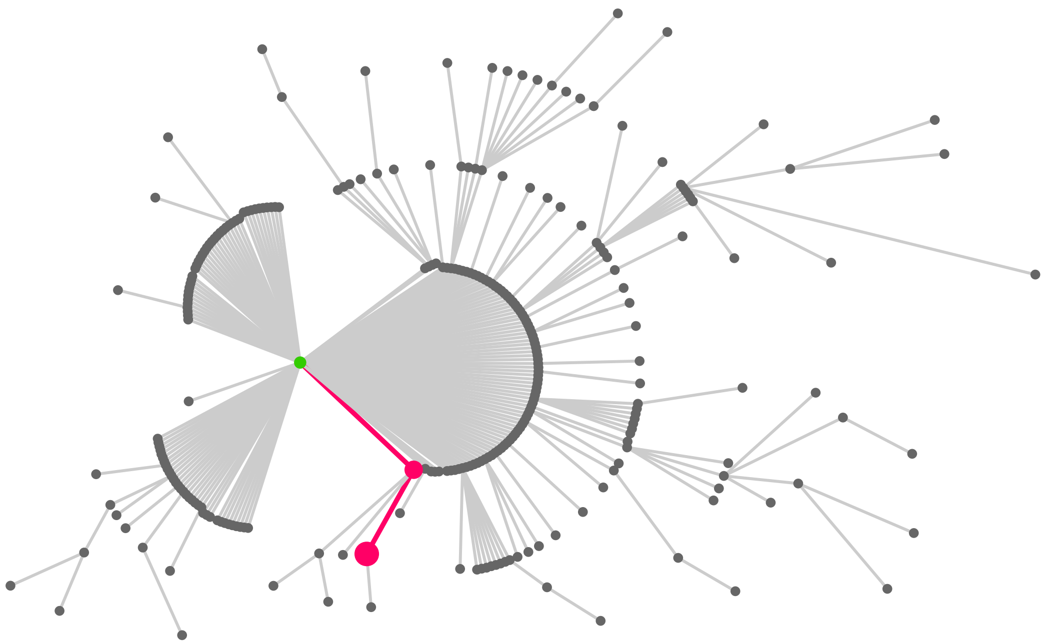
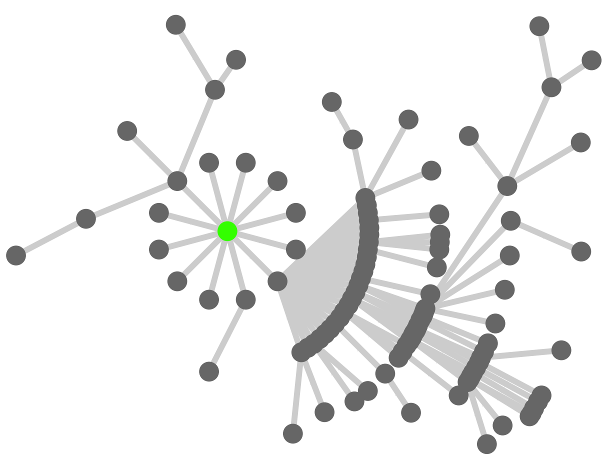


Figure S4: Tweet count over time of MIT red balloon team versus another MIT-related event during the same month. For easy comparison, we shifted data temporally by matching the day when the MIT team launched the online campaign with the day when the MIT bike news was released: (a) Daily Twitter counts for both events (Data are scaled in this figure so that both peaks have the same value); (b) Raw daily increase in Twitter counts. The vertical blue dash line indicates the day of the DARPA Challenge competition.

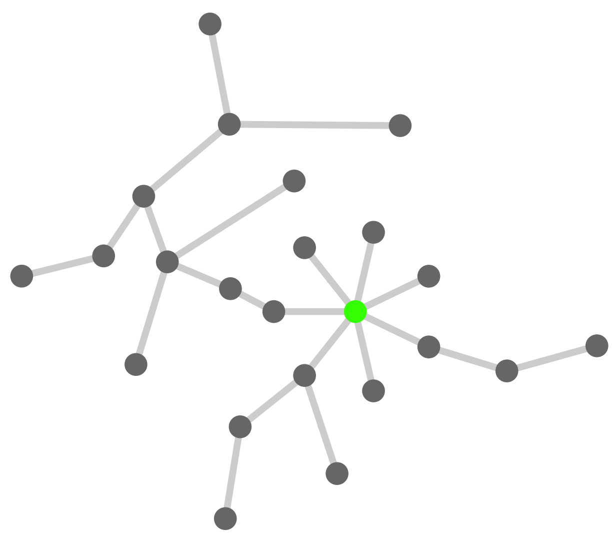
6 Examples of recruitment trees



(a) Large successful recruitment tree



(b)



(c)

Figure S5: (a) A tree with the root is shown in green, and the successful path highlighted in red. (b) and (c) Two additional networks that did not lead to balloons.