

The Rippling Effect of Social Influence via Phone Communication Network



Yan Leng, Xiaowen Dong, Esteban Moro, and Alex ‘Sandy’ Pentland

1 Introduction

We live in a connected world and are increasingly closer to each other thanks to the emerging information technologies. While the “small-world” phenomenon and the “six degrees of separation” have been traditionally studied by Milgram [14] and Watts [24], a recent research suggests that the average degree of separation between two members of the online social network Facebook is reduced to around 4.74 [5]. Furthermore, individuals are not merely connected; as a series of experiments in various domains such as obesity, happiness, cooperation, and political opinions has demonstrated, connectivity also indicates behavioral similarities of up to three degrees of separation [7, 25].

The recent availability of large-scale communication and networked data, such as emails, mobile phone records, and online social media activities, enables the studies of information diffusion and correlations of adoption behaviors as well as social contagion processes at an unprecedented scale [8, 15, 17]. In particular, the understanding of the phenomenon of and the mechanism that drives the social contagion process help promote behavioral change in domains such as commerce, public health, politics, and social mobilization at both local and global scales [3, 6, 9, 23]. As examples, Aral et al. [4] focused on the diffusion of the adoptions of mobile service application using a social network connected by instant message

Y. Leng · X. Dong · A. ‘Sandy’ Pentland (✉)
MIT Media Lab, Cambridge, MA, USA
e-mail: yleng@mit.edu; xdong@mit.edu; pentland@mit.edu

E. Moro
MIT Media Lab, Cambridge, MA, USA

Departamento de Matematicas & GISC, Universidad Carlos III de Madrid, Leganes, Spain
e-mail: emoro@mit.edu

traffic [2]. Ugander et al. [23] found that the decision to join Facebook varies with the number of distinct social groups their friends occupy. Bond et al. [6] conducted a 61-million-person experiment on Facebook and found that strong ties are instrumental for spreading political behavior through online social network. Often, the connectivity and structure of the social network play a role in the effectiveness of social contagion. For instance, both Onnela et al. [17] and Ugander et al. [23] emphasized the importance of network structure in information spreading and product adoption [26].

However, most of the previous works focus on online social networks, and measure influence between direct contacts concerning either long-term habits or low-cost decision-making in virtual space (such as online product adoption). In this study, we are interested in investigating how social influence propagates over a large-scale offline communication network, and how it manifests in short-time decision-making and social mobilization that are more costly than merely information diffusion or online production adoption.

We use a data set of mobile phone records with high resolution in Andorra for our analysis. We construct a large-scale communication network and mirror the contagion process of social influence, whose effect is measured by the change in the likelihood of attending a large-scale international cultural event in the capital city. In order to control for the selection bias caused by homophily and identify the causal effect of social influence, we utilize a matching method to mimic the procedure of random assignment of treatments [3, 11]. One novel aspect of our study is to condition matchings on revealed preferences, i.e., historical visitation patterns, instead of the traditionally considered demographics. Rather surprisingly, our results show that influence decays across social distance from initial attendees, but persists up to six degrees of separation, similarly to the physical phenomenon of ripples expanding across the water. Meanwhile, the patterns of communication, such as intensity and the timeliness of communication, also impact the strength of social influence, but to a lesser degree. Finally, we analyze the heterogeneous effects of social influence on the population, and observe that the effect is stronger on the geographically explorative subgroup of population.

2 Data and Method

Mobile phone logs have been used in various studies as a proxy for human mobility and social interactions at a societal scale [8, 21]. We leverage the detailed tracking and wide coverage of mobile phone logs in the country of Andorra to study how the likelihood of an individual attending a local Cirque Du Soleil performance, which was held repetitively in July, 2016, is affected if someone in his social circle receives phone calls directly or indirectly from past attendees of the event.

We introduce three key definitions in our study. First, we assume that people who were connected to a cell tower nearby the performance venue, as shown in the left panel of Fig. 1, during the performance hours (± 30 min as buffer time)

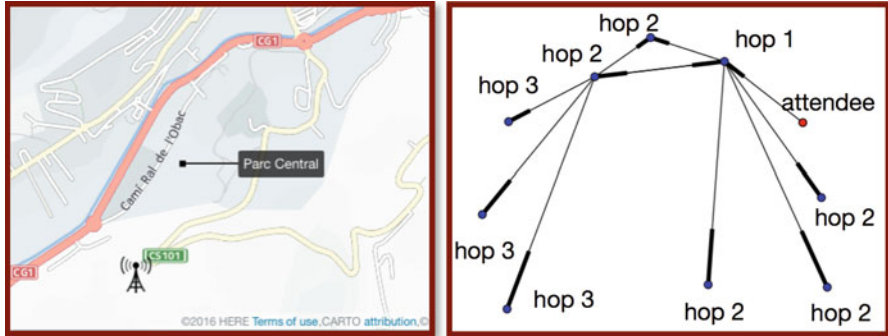


Fig. 1 Illustration of attendee, influence cascade and hop

attended the events and are labeled as **attendees**. Next, we construct **influence cascade**, as shown in the right panel of Fig. 1, by adding links between the caller and receiver if: (1) at least one of them is linked directly or indirectly with the attendees by the time the call was initiated; (2) the calls took place within 24 h after the performance started. Finally, we use **hop** to capture the shortest social distance to any attendee via the influence cascade. Overall, we observed 16,043 attendees across the one-month observational period. Among others, the influence cascade covers 161,857 individuals. And another 71,337 population are disconnected to the influence cascade.

In order to quantify the effect of social influence in people’s decision-making, the key challenge is to control for the upward estimation bias caused by homophily. We use matched sample estimation to mimic the assignment of treatment as in a randomized experiment, rather than regression analysis which only establishes correlations [3, 10, 11]. More specifically, for the influence cascade constructed for each day, we consider a treatment group in which individuals are of certain social distance from the attendees (we use treatment group on hop h to represent people that are h -degree of separation from the closest attendee), and a control group in which individuals are not connected to any attendee on that day. Individuals in treatment and control groups are matched to control group on a one-to-one basis based on their mobility patterns, which we will further explain in more detail in later section.

Before establishing causal studies, we first analyze the distribution of social distances of individuals to the attendees. As shown in Fig. 2, a large mass of population are three and four degrees of separation from the attendees. Moreover, we analyze the predictive power of the degree of separation from the attendees for attendance rate. We compare the attendance ratio between treatment group on hop h and control group. The larger-than-one ratio comes from a mixed effects of homophily, social influence and other confounding variables. The right panel of Fig. 2 shows the ratio between the likelihood of attending the social event of people on hop h and those who receive no treatment. As we see, direct contacts of the

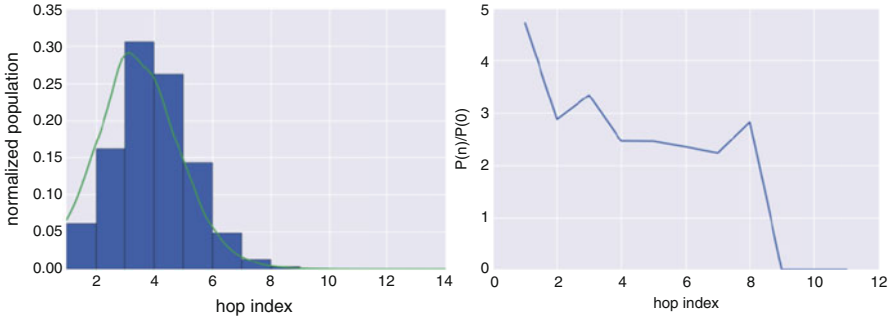


Fig. 2 Distribution of social distances of individuals to the attendees (left) and the attendance ratio between treatment group on hop h and control group (right)

attendees are five times more likely to attend the performance than individuals who do not receive treatment. Meanwhile, individuals on hop six are 2.5 times more likely to attend the events than individuals receiving no treatment. The average decreasing trend of the likelihood indicates that the degree of separation from the attendees is an important factor in studying the likelihood of attending the performance. However, this correlation does not indicate causality, the latter of which is the main focus of our study.

2.1 Controlling for Homophily

It is widely argued that the adoption behavior in the social network (the decision of attending the event in our case) is a mixture of similarities over friends and contagion driven by social influence [3, 6, 22]. Similarities among peers may cause the over-estimation of social influence [3]. Therefore, we need to balance the distribution of similarities across individuals in the influence cascades and isolate the causal effect of word-of-mouth influence through phone calls in our observational study.

Empowered by the longitudinal and detailed mobility tracking via Call Detail Records, we use behavioral patterns to characterize individuals instead of the widely applied method of demographic characterization [3, 6]. The power of behavioral characterization as a control for homophily is that behavior reveals preferences regarding the same type activities that we are observing and treating [13], which is exactly what we want to control for. Specifically in our case, activities performed during their leisure time, the revealed visitation preferences, are captured via cell tower visitation frequencies over the weekend for the past 6 months [12]. As shown in Fig. 3, the left panel represents an individual who spends most of the weekends in the crowded shopping districts while the right panel stands for an individual with a diversified activity patterns.

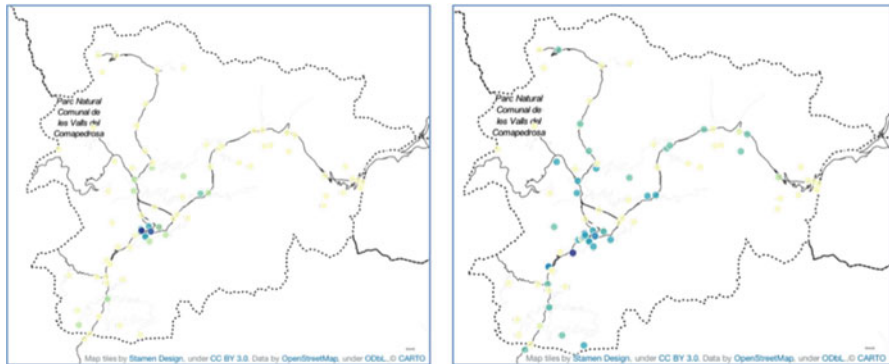


Fig. 3 Two examples of historical mobility patterns during the weekends for the past 6 months

2.2 Matching

As stated before, we use matched sample estimation to yield the estimates of social influence by conditioning matches on mobility frequency vectors. The matching results establish an upper bound to which extent social influence, rather than homophily, explains the attendance behavior¹ [3].

We segment individuals into two groups, the treatment group and the control group, based on whether they receive influence related to the event or not. Treatment groups are further split into eight subgroups according to the hop index. The control group consists of individuals who are disconnected to the influence cascades. Each individual in the treatment group is paired with another individual in the control group that is most similar in terms of preferences approximated by mobility patterns. By such a matching, we ensure that the main difference between the two individuals paired together is whether or not one receives the treatment of social influence [20]. The matchings depend on nearest Mahalanobis distance calculated as:

$$\text{md}(X_j, X_k) = [(X_j - X_k)^T S^{-1} (X_j - X_k)]^{1/2}, \quad (1)$$

where X_j and X_k are the covariate vectors (mobility frequency vectors) for individual j and individual k , and S is the sample covariance matrix for the mobility frequency matrix X .

We perform Principal Component Analysis on X to reduce the correlations of the visitation patterns among nearby cell towers and to reduce the number of variables used in matching. Dimension reduction is important in Mahalanobis Distance Matching, which works better in balancing fewer covariates [11].

¹Unobserved confounding variables are difficult to control for by using matching-based methods. To partly address the issue that tourists may travel together and social links may not pass social influence, we remove individual pairs who are potentially on the same trip to Andorra. This can be inferred based on whether individuals stay at the same hotel at the same night.

In our setting, the difference in the attendance rate of the two groups is the average treatment effect of social influence:

$$ATE_h = E(Y_{ih} - Y_{ic}), \quad (2)$$

where ATE_h is the average treatment effect of treatment group on hop h , Y_{ih} is the outcome for matched pair i in treatment group h , and Y_{ic} is the outcome for matched pair i in control group.

3 Results

In this section, we first investigate the effect of social influence after distinguishing it from homophily using Mahalanobis Distance Matching. To evaluate the inflation bias caused by homophily, we compare our results with random matching, where we do not control for homophilous behavior and pair individuals randomly. Furthermore, we quantify both external and internal factors that affect the strength of social influence, namely, the patterns of the communications and the characteristics of the individuals.

3.1 *The Decay of Influence over Social Distance from Attendee*

After distinguishing homophily and social influence, we are able to estimate the treatment effect of social influence on the likelihood of attendance. In Fig. 4, the blue-dashed line shows the average treatment effect of social influence (as in y -axis) across hops (as in x -axis). The positive treatment effects—the increasing likelihood of attending a future performance—indicate that social influence promotes the likelihood of attending the performance. More importantly, we discover a “ripple effect” of social influence over communication network: originating from the attendees and expanding across information cascade. In particular, this effect decays across social distances from the attendees and persists up to six degrees of separation. The average treatment effect of social influence is 11% on the first hop and drops dramatically to a half at the second hop. Starting from the third hop, the treatment effects decay slowly and persist until the sixth hops.

The difference between the red-dashed line and the blue-dashed line in Fig. 4 shows the overestimation of social influence without controlling for homophily. In particular, with random matching, we overestimate the effect of social influence by around 100%, which is similar to the findings in a previous study by Aral (2009) on the adoption of an online application [4].

Furthermore, we use “random shuffling” proposed by Anagnostopoulos [1] to exclude the concern that other mechanical reasons might cause the decay pattern in

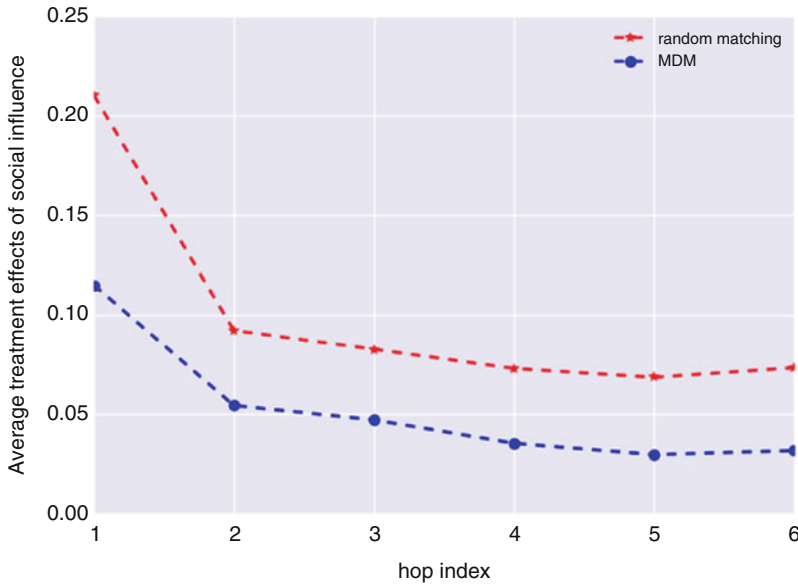


Fig. 4 Average treatment effect of social influence via communication network

social network. We first randomly assign people to control and treatment group, as well as the hop index if assigned to treatment group, and then measure the average treatment effect with Mahalanobis Distance Matching. The average treatment effect as well as the decay pattern disappear.

In order for the estimation of treatment effects from matching results to be robust, the assignment of treatment, conditional on the Mahalanobis distance, need to be as good as randomly assigned. In other words, the covariates are required to be balanced between matched pairs in treatment and control groups. Therefore, we use standardized mean differences (SMD) to evaluate whether the covariates in the two groups demonstrate sufficient overlap [16]. SMD is calculated as the difference of means in units of pooled standard deviation as follows:

$$SMD = \frac{\bar{x}_{l,h} - \bar{x}_{l,c}}{\sqrt{(s_{l,h}^2 + s_{l,c}^2) / 2}}, \tag{3}$$

where $\bar{x}_{l,h}$ and $\bar{x}_{l,c}$ are the means of covariate x_l for treatment group h and control group, respectively, and $s_{i,h}$ and $s_{i,c}$ are the standard deviation of covariate x_l for treatment group h and control group, respectively. We run the covariates balanced test and show that all of the SMDs are far below 0.1, which rejects the hypothesis that they have insufficient overlap.

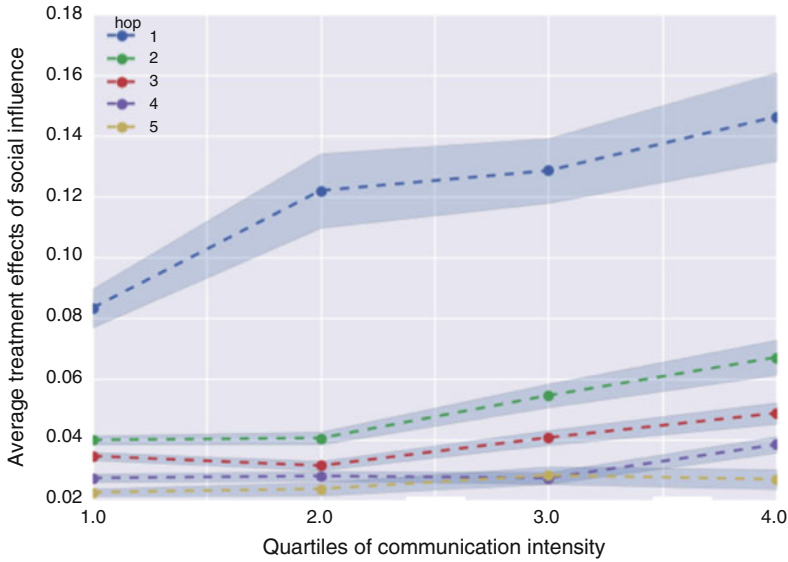


Fig. 5 Average treatment effect of social influence with respect to intensity of communications. Different colors shown in the legend represent different hop indexes as labeled

3.2 Communication Patterns

In this section, we test the hypothesis that social influence and contagion process on the social network may vary according to the communication patterns. As shown in Fig. 5, more intense communications between two individuals indicates a larger treatment effect for the first three hops and stay constant afterwards. In terms of the timeliness of communication, we show in Fig. 6 that the treatment effects are significantly stronger if the calls are made immediately after the event. Similarly to intensity, this only holds up to hop three. These two empirical exercises indicate that communication patterns exert quantifiable and discernible effects on the strength of social influence up to three degrees of separation.

4 Discussion

In this study, we illustrate the application of a matching strategy in a large-population study to identify the effect of social influence. A novel aspect of our study is the use of matched samples as determined by previously observed behavior instead of those obtained by Randomized Control Trials (RCTs), which seems potentially quite useful in many large-scale studies. By analyzing the pattern of attendance of an international cultural event in Andorra using large-scale mobile

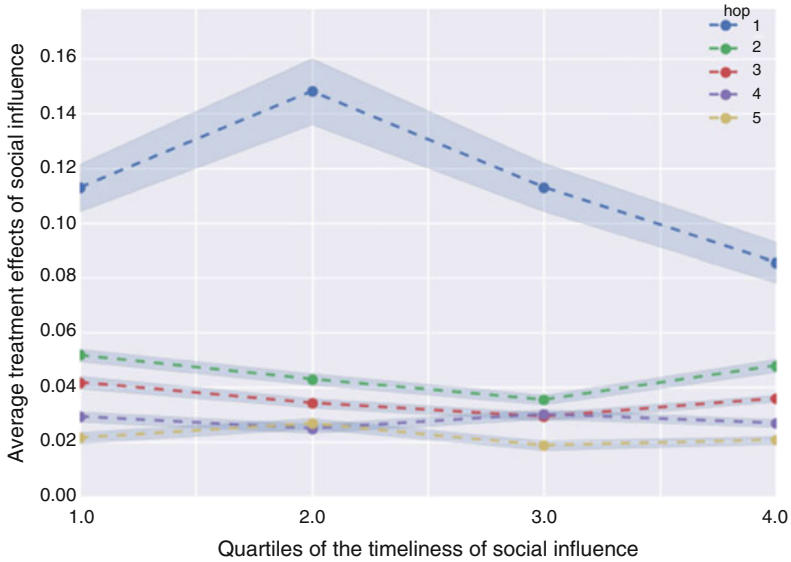


Fig. 6 Average treatment effect of social influence with respect to timeliness of communication. Different colors shown in the legend stand for different hop indexes as labeled in the legend

phone data, we quantify how our decision-makings are influenced by, and how the social network propagates our influence to, people that are several degrees away from us in the communication network with matched and balanced samples.

Our results reveal the subtle and often invisible effect of social influence on decision-making via phone communication network, which, surprisingly, persists up to six degrees of separation. This is analogous to the physical phenomenon of ripples expanding across the water, which highlights the hidden relationship and connections among people in the society. More interestingly, we show that such effect is significantly larger when phone communication took place immediately after the event and lasted longer, and when those receiving calls are more explorative geographically as indicated by a more diverse mobility pattern.

The ripple effect via phone communications demonstrated through our study has far-reaching implications in domains such as viral marketing, public health, and social mobilization. Recent works have demonstrated the success of social mobilization via Internet-based services [18], but also shown that such mechanisms are not without limitations [19]. Our findings suggest that an alternative would be to exploit the hidden and often overlooked influence between people that are caused by chains of offline communication. The same strategy may also be applied into marketing or political campaigns. Our results on the impact of communication pattern and mobility pattern of individuals on the strength of influence can also help design more effective strategies to maximize social influence.

Our work also opens new possibilities in understanding social influence and contagion, in terms of both mathematical modeling and experiment-based studies.

In the context of networks, threshold-based contagion models and epidemic models have largely explored the direct interaction between neighboring nodes in the network, where the behavior of a given node is dependent on its interactions with neighboring nodes. Hidden interactions across several degrees of separation could be naturally incorporated into such models. For example, we could systematically model the treatment effect and the adoption behavior of a given node as a function of degrees of separation, as well as other network characteristics. With better models on contagion processes, we could perform counter-factual simulations over different intervention strategies to incentivize key individuals and maximize social influence for behavioral change.

It is worth noting that our study also has certain limitations. First, given that we do not have the actual records for attendance of the event, we consider people who had phone activities at cell towers close to the venue as attendees. This strategy might, therefore, have included people who just passed by the venue without actually attending the event. Second, due to the lack of demographic information, we approximate homophily in a social network by looking at the mobility history of individuals. While it is reasonable to assume that mobility patterns reflect to some extent characteristics and interests of different people, it may also make people with different demographics much more similar. Third, in the current framework, we define social distance as the length of the shortest path between an individual and the attendees, thus effectively considering only this “strongest treatment” in estimating the treatment effect. There might be a multiplicative effect in the case of more than one communication path (hence the possibility of multiple treatments), which may require slightly more complex modeling of influence. We leave such analysis for future work.

References

1. Anagnostopoulos A, Kumar R, Mahdian M (2008) Influence and correlation in social networks. In: 14th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 7–15
2. Aral S (2011) Commentary-identifying social influence: a comment on opinion leadership and social contagion in new product diffusion. *Mark Sci* 30(2):217–223
3. Aral S (2012) Social science: poked to vote. *Nature* 489(7415):212–214
4. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci* 106(51):21544–21549
5. Backstrom L, Boldi P, Rosa M, Ugander J, Vigna S (2012) Four degrees of separation in Proceedings of the 4th annual ACM web science conference. ACM, New York, pp 33–42
6. Bond RM et al (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415):295–298.
7. Christakis N, Fowler J (2009) *Connected: the surprising power of our social networks and how they shape our lives*. Little Brown and Company, New York
8. Deville P et al (2016) Scaling identity connects human mobility and social interactions. *Proc Natl Acad Sci* 113(26):201525443
9. Gleeson JP, Cellai D, Onnela JP, Porter MA, Reed-Tsochas F (2014) A simple generative model of collective online behavior. *Proc Natl Acad Sci* 111(29):10411–10415

10. Hill S, Provost F, Volinsky C (2006) Network-based marketing: Identifying likely adopters via consumer networks. *Stat Sci* 21(2):256–276
11. King G, Nielsen R (2016) Why propensity scores should not be used for matching
12. Leng Y, Rudolph L, Pentland A, Zhao J, Koutsopolous HN (2016) Managing travel demand: location recommendation for system efficiency based on mobile phone data. *CoRR abs/1610.06825*
13. Lobel I, Sadler E (2015) Preferences, homophily, and social learning. *Oper Res* 64(3):564–584
14. Milgram S (1967) The small world problem. *Psychol Today* 1(1):61–67
15. Miritello G, Moro E, Lara R (2011) Dynamical strength of social ties in information spreading. *Phys Rev E* 83(4):045102
16. Normand SLT et al (2001) Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *J Clin Epidemiol* 54(4):387–398
17. Onnela JP et al (2007) Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci* 104(18):7332–7336
18. Pickard G et al (2011) Time-critical social mobilization. *Science* 334:509–512
19. Rutherford A et al (2013) Limits of social mobilization. *Proc Natl Acad Sci* 110(16):6281–6286
20. Stuart EA (2010) Matching methods for causal inference: a review and a look forward. *Stat Sci Rev J Inst Math Stat* 25(1):1
21. Toole JL, Herrera-Yaqüe C, Schneider CM, González MC (2015) Coupling human mobility and social ties. *J R Soc Interface* 12(105):20141128
22. Toulis P, Kao EK (2013) Estimation of causal peer influence effects. *ICML* 28(3):1489–1497
23. Ugander J, Backstrom L, Marlow C, Kleinberg J (2012) Structural diversity in social contagion. *Proc Natl Acad Sci* 109(16):5962–5966
24. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684):440–442
25. VanderWeele TJ (2013) Inference for influence over multiple degrees of separation on a social network. *Stat Med* 32(4):591–596
26. Zhao J, Wu J, Xu K (2010) Weak ties: Subtle role of information diffusion in online social networks. *Phys Rev E* 82(1):016105