Modelling the distance impedance of protest attendance

V.A. Traag[a,b,c], R. Quax[d], P.M.A. Sloot[d,e,1]

[a] *KITLV, Leiden, the Netherlands*
[b] *e-Humanities Group, KNAW, Amsterdam, the Netherlands*
[c] *Centre for Science and Technology Studies, Leiden University, Leiden, the Netherlands*
[d] *Faculty of Science, University of Amsterdam, Amsterdam, the Netherlands*
[e] *Research University ITMO, St. Petersburg, Russia*
[f] *Complexity Institute, NTU, Singapore*

Abstract

Protesters are usually young, relatively well educated, middle class people that are politically engaged. But where do protesters come from? We here show, based on mobile phone data, that distance is an important impedance to protest attendance. Most protesters come from nearby regions, suggesting distance forms an obstacle to participation. Although this effect can be partly explained by social network effects, which show similar spatial dependencies, an effect of distance remains. This suggests distance still acts as an obstacle to participation, although it may also be that long-range contacts are less effective for recruitment. Face-to-face contacts seem more important in spreading protests through earlier participants, whereas central recruitment works better by telephone. Our results are important for understanding processes of recruitment.

Keywords:

1. Introduction

In recent years many protests have captured the attention of people across the globe. In 2011 protests broke out across the Arab world, aspiring to a more democratic rule. In early 2014 protests in Kiev led to the fall of Yanukovich. The economic crisis triggered protests ranging from the Spanish Indignados to Occupy Wall Street. From earlier studies, we know that protesters are generally young, well educated, biographically available, and politically engaged [1]. But where do protesters come from?

Email address: v.a.traag@cwts.leidenuniv.nl (V.A. Traag)

Protests and contentious politics more generally have been intensely studied [2, 3, 4] and scholars are long fascinated by how they diffuse [5, 6]. Most of the literature focuses on large-scale diffusion of protests [4, 5] rather than the diffusion of individual participation in protests. As such, they analyse how the protest as a whole diffuses, not how individual participation comes about. Individual participation is more closely studied in the context of mobilisation and recruitment [7]. Under the umbrella term of "differential recruitment" scholars analysed what type of people tend to participate in protests. Besides the aforementioned individual characteristics, one of the most clear empirical regularities in studies of protest recruitment is the role of the social network [7, 8, 9, 10, 11, 12, 13]. There are usually many people that are positively disposed towards the goals of a protest [8]. Yet few participate unless they are asked by others [9]. Knowing people in the movement is one of the best predictors of participation [10], especially in high-risk activities [11].

Yet this literature on recruitment does not address how distance impedes participation to protests. We address that caveat here, using mobile phone data. We find empirically that distance indeed affects participation and that most protesters have local origins. In addition, we model this empirical regularity using various approaches. Although the models considered may be somewhat rudimentary, they serve as a first stepping stone towards more realistic models. Social networks are known to have a strong impact on participation, but also exhibit spatial regularities [14, 15]. We find that although the social network partly explains the spatial pattern of protest participation, distance still has an effect. This suggests that distance acts as an obstacle to participation.

2. Results

2.1. From mobile phones to protests

We analyse protest participation in an undisclosed African country on the basis of mobile phone data. Such data have been used before to study attendance of people in other events, such as baseball games or theatre shows [16] or simply their presence at mass gatherings [17] and to study reactions to emergencies [18, 19]. Yet protests have never been analysed using mobile phone data. Such mass gatherings can be detected based on mobile phones [20, 21], although it is difficult to determine the exact type of event. Mobile phone data also provide insight in human mobility in general, and reveal lots of regularities [22, 23, 24, 25, 26, 27, 28, 29], see [30] for an extensive review. Protests may take many different forms [3], but we

restrict ourselves here to those forms that require a physical presence, thus excluding forms such as petitions or letter writing. Other research has relied on online service such as social media. Twitter has for example been used to study the attendance of protests [31] while Flickr has been used to study the occurrence of protests [32]. The penetration rate of such services is quite low in African countries [33], thus precluding their use for our study.

We estimate participation in a protest by considering the location and time information from the mobile phones. The idea (similar to [20]) is that we assume people to participate in a protest if they are present during the protest, but otherwise not regularly there (see method section for details). Obviously, this will entail a certain degree of error: (1) some people will participate, but not show up on the mobile phone data; (2) not all people that are present will participate in the protest; and (3) some people that are regularly present in that location can still protest. Nonetheless, it was shown earlier that this technique can uncover participants [20]. In general, mobile phones are expected to show some bias related to socio-economic (wealthier people more likely to own a mobile phone) and geographic (urban people more likely to own a mobile phone) conditions. We expect the errors in our methodology for identifying participation to increase the uncertainty, but not to produce any additional bias.

We consider users to reside at the location where they are most regularly seen (see method section for details). This way we infer the number of protesters m_j and the total number of users n_j from a location j. In addition, we know the number of calls w_{ij}^{calls} and the total duration of the calls w_{ij}^{duration} between two locations i and j. Based on the individual level CDR data we also estimate the total number of trips w_{ij}^{trips} between two locations i and j (see method section for details).

We used the Armed Conflict Location & Event Data (ACLED) project[1] for determining where and when protests took place [34]. They recorded political violence and conflict from 1997 onwards in over 50 different countries and provide latitude-longitude coordinates of all events. We thus only use the ACLED data to determine where and when protests took place. We study 39 events marked as "riots/protests" from a single year for which we have mobile phone data. They took place throughout the year and were not particularly clustered in time. The events have been covered by local media and press agencies, and at time by international agencies, but none of them received much attention in the international media. We analyse

[1] http://www.acleddata.com/

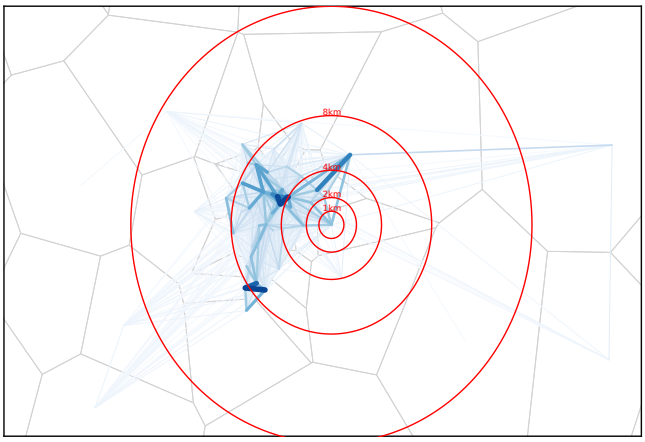


Figure 1: **Protest participation.** Protesters usually concentrate around the protest. The distance dependence can be well captured by geographical models such as the gravity model (as illustrated by the red circles). The spatial patterns may be partially explained by the communication network (as indicated by the lines in blue). If a location is better connected to the protest location, there are generally more protesters from that location. Even when controlling for the social network, distance has an effect on participation. Long distance communication is thus less effective for recruiting people, or remains an obstacle for people to participate.

the geographic origins of people present during these events. We infer the presence of people from the usage of mobile phones in the antenna that is closest to the coordinates provided by ACLED. The density of such antennas is typically higher in urban areas than in rural areas, leading to a resolution of several hundreds of meters in the former, but a resolution of several (dozen) kilometres in the latter. Of the 39 events, we found 32 that took place in an urban environment, while 7 took place in a rural setting (including small towns up to 20 000 inhabitants). Similarly, we infer the origin of people based on the most frequently used antenna. Using this method, we uncovered around 300–400 participants in our dataset (average 340) for the 39 protests, but it varied from 27 to 1105 (Fig. 2). On average, the urban protests had 394 participants coming from 15 km away, while the rural protests only had 92 participants coming from 33 km away. It could be that this is affected by a bias in mobile phone ownership, which is usually concentrated in urban areas. The lower antenna density in rural areas may also bias these results, as antennas are further apart than in urban areas. Unfortunately, no alternative source is available to corroborate our estimate

of the number of participants. Additionally, some other events may have taken place simultaneously, thus obfuscating our results. Nonetheless, it is unlikely this is the case for all 39 protests. See the methods section at the end for more details.

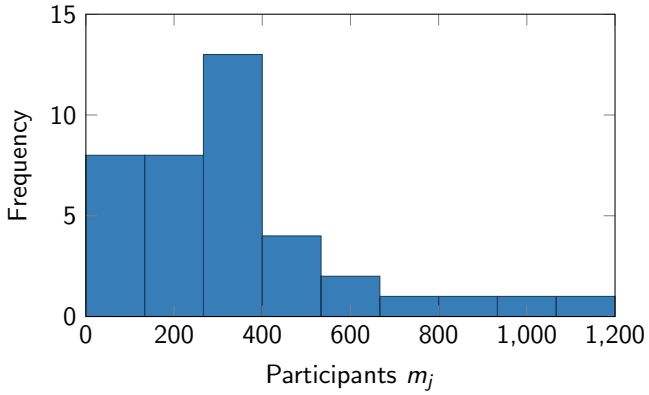


Figure 2: **Number of participants.** Histogram of the number of participants uncovered by the mobile phone data for the different protests. This is concentrated around 300–400 participants for most protests.

2.2. Analysis

We find that most protesters indeed come from relatively close by. The average distance between the origin of a protester and the protest location is about 20 km. On average, roughly 83% of the protesters come from within the average distance. But how does participation depend exactly on the distance to the protest location? There are various factors that may influence how protest attendance decays over distance (see Fig. 1 for an illustration). Participation may simply decrease because it cost something to overcome a certain distance. But there may also be some network effects: people further apart are less likely to be connected.

2.2.1. Geographic models

We initially model distance as an impedance to participation. Let us denote by i the location of the protest, and we are then interested in the number of participants m_j from location j. Following classical geographical models [35], the impedance increases with distance d_{ij} between location i and j. If the cost to participate increases linearly with distance, we obtain an exponential decay of participation with distance, whereas if the cost depends

logarithmically on the distance we obtain a powerlaw decay [36, 37]. This behaviour could be derived from maximum entropy arguments [37], although the foundations of the gravity model have been debated [26]. Additionally, the number of participants from location j scales with the number of people n_j, where we allow for non-linear scaling. Denoting by $\hat{m}_j$ the predicted number of participants, we thus obtain the classical gravity model[2]

$$\hat{m}_j = Z \frac{n_j^{\beta}}{(d_{ij} + \delta)^{\gamma}}, \tag{1}$$

where Z, β, γ and δ are free parameters that are estimated from the data. Higher β means that relatively more people come from more denser populated areas. We here use the powerlaw decay $(d_{ij} + \delta)^{\gamma}$ which works better than exponential decay and a powerlaw decay with exponential cut-off [35] (see method section for fitting details). In general, higher γ increases the decay of participation over distance. For $\gamma = 0$ there is no dependence on the distance, and for $\gamma = 1$ the decay is linear (i.e. half the people participate at twice the distance). We use $(d_{ij} + \delta)^{\gamma}$ instead of d_{ij}^{γ} since we also estimate the number of protesters at the protest site itself, for which $d_{ij} = 0$. We find that this model quite accurately predicts the number of protesters from a location j ($R^2 \approx 0.63$, for alternative fitting statistics, see Table A.1 and methods[3]). The number of participants scale with distance with an exponent of $\gamma \approx 1.8$, and with the number of users with exponent $\beta \approx 0.78$ (Fig. 3(a)–(d) and Table A.2). Every doubling of the distance thus reduces the number of participants to roughly one third. The scaling with the number of users is sublinear, which surprisingly suggests that more populous locations are less likely to engage in protest, whereas more urban participation would be expected [38].

Arguably, the effect of distance is mostly an effect of overcoming obstacles to participation [8]. Travelling a certain distance always costs money and time, and as such, it may become increasingly unlikely to exceed any (perceived) benefits from the protest. However, it has been argued that rather than distance d_{ij} itself, the number of intervening opportunities in

[2]The classical gravity model stipulates that the interaction between two locations is a product of both n_i and n_j. We estimate the participation for each protest separately, so that n_i is constant for that protest.

[3]Generally, R^2 is flawed for model selection in the case of non-linear least squares. We therefore corroborate our findings using alternative statistics such as Akaike's Information Criterion (AIC) and the Root Mean Square Error (RMSE). The alternative statistics are in line with the conclusions drawn from R^2.

between two locations is more important in explaining spatial patterns [39]. From the perspective of protesters, the idea of (intervening) opportunities makes no sense. After all, it is unlikely that one finds alternative protests closer by (unless there are nation wide protests, which would make an interesting case). However, from the point of view of recruitment there are many opportunities. After all, to a recruiter, everybody is a potential protester and thus presents an opportunity for recruitment. Let us assume that the organisers are trying to recruit people from the protest location. Generally, recruiters try to maximize their impact with minimum means [40]. Since travelling takes time and effort, it is more efficient to recruit close by. Moreover, people close by may be more responsive, so that it may also be more effective. Hence, the organisers first try to recruit people close by, and only try to recruit people further away whenever they could not recruit anybody closer by. For simplicity sake, let us assume that everybody has the same fixed recruitment probability p. So, whenever somebody is contacted by the organiser, with probability p he or she will join the protest. Somebody from location j will only be contacted by a recruiter if nobody closer by joined. The probability nobody joined before is $(1-p)^{n_i+n_{ij}}$, where

$$n_{ij} = \sum_{\substack{k \neq i \\ d_{ik} < d_{ij}}} n_k$$

denotes the number of people closer to location i than location j. The probability that somebody in location j did join is then $1-(1-p)^{n_j}$. Assuming there are Z (successful) recruiters, the total number of participants from location j is then expected to be

$$\hat{m}_j = Z(1-p)^{n_i+n_{ij}}(1-(1-p)^{n_j}). \tag{2}$$

We estimate Z and p from the data for each protest, while n_i and n_{ij} are directly available. Interestingly, this intervening recruitment model is identical to an early model of intervening opportunities [41], which has become a standard formulation in geography [35]. This intervening recruitment model performs much better than the recently introduced radiation model [26] ($R^2 \approx 0.55$ versus $R^2 \approx 0.31$). It fits the data less well than the gravity model though (see Fig. 3(a)–(d) for comparing the two). This model of recruitment is thus less likely, and participation drops more because of pure distance than because of the number of people closer by. As such, either the model of recruitment is unrealistic, or distance itself functions to create an obstacle to participation [8], independently of the population density.

2.2.2. Network models

When people join a protest, they often already know others in the movement [7]. Social networks are known to exhibit spatial dependencies that can be modelled by the gravity model [14, 15] or the radiation model [26]. Our social network only consists of interaction between locations, and we try to model this interaction w_{ij} between two locations i and j. We find that the gravity model provides a good fit ($R^2 \approx 0.67$) for both the communication and the mobility in our dataset (Fig. 4 and Table A.3). Both scale approximately as $w_{ij} \sim \frac{\sqrt{n_i n_j}}{(d_{ij}+\delta)^\gamma}$, with $\gamma \approx 1.18$ for the number of calls and the number of trips, while the duration decays much slower with $\gamma \approx 0.86$. Nonetheless, the gravity model underestimates the interaction at larger distances (see Fig. 4(a)–(c)). The radiation model is again a significantly poorer fit ($R^2 \approx 0.54$). In short, the social network also shows a clear distance effect, and perhaps decreasing participation is simply an effect of the social network. We now investigate this possibility in this section.

We again assume that the organisers are trying to recruit people from the protest location. In this network model we assume they will call people. The total number of calls in between two location is w_{ij}^{calls}, so that the average number of calls somebody in location j receives from i is $\Lambda_{ij} = w_{ij}^{\text{calls}}/n_j$. We assume again that everybody has a fixed recruitment probability p for each call. The probability that somebody from location j will join the protest is then $1-(1-p)^{\Lambda_{ij}}$ if he had been contacted by the organisation. Although it could be argued that the contact intensity Λ_{ij} should be scaled with a parameter α to arrive at $1-(1-p)^{\alpha\Lambda_{ij}}$, this is a degenerate formulation, as we can also write this as $1-e^{-\alpha\phi\Lambda_{ij}}$ with $\phi = -\log(1-p)$, so that there is essentially only a single parameter $\gamma = \alpha\phi$. In total there are n_j individuals at location j, and we assume only a fraction of q have been contacted by the organisation. Similar to the gravity model, we allow for a non-linear scaling with the number of users, and arrive at

$$\hat{m}_j = q n_j^\beta (1-(1-p)^{\Lambda_{ij}}). \tag{3}$$

We call this the direct contact model, because we do not consider any additional cascades of participation or second order effects. We find that this model fits well ($R^2 \approx 0.68$) using the number of calls (see Fig. 3(e)–(h)), although the mobility also works quite well ($R^2 \approx 0.67$), but the duration works less well ($R^2 \approx 0.58$). The scaling with the number of users varies and is superlinear for communication but sublinear for mobility, so that results on this effect remain inconclusive.

It could be argued that rather than people being recruited only by the or-

ganisation, participants may also recruit others to join the protest. Hence, there may be additional cascades of participation or indirect, higher order effects, thus triggering further diffusion. We here use the so-called SI model [42] in which every protester has a fixed probability of convincing a friend or acquaintance to join the protest. However, we do not dispose of the individual network, only of the aggregated communication and mobility network between antennas based on the mobile phone dataset (see methods for details). We therefore assume that participants recruit people in the same location at a constant rate γ. Additionally, protesters recruit people in other locations at a rate proportional to Λ_{ij}. Taking this approach, we arrive at the dynamical system

$$\dot{m}_j = \left[\beta \sum_i \Lambda_{ij} \frac{m_i}{n_i} + \gamma \frac{m_j}{n_j}\right] (m_j - n_j). \tag{4}$$

This type of model is known as a metapopulation model, because there are both internal dynamics in each location as well as dynamics across locations [43]. In order to fit this model, the initial condition is set to a single protester in the protest location. We solve the dynamics numerically for a fixed time $t = 1$ since alternative times simply correspond to a rescaling of the parameters β and γ. The epidemic model provides quite a good fit (see Fig. 3(i)–(l)), so that indeed a diffusion process might take place on the network. The best fit ($R^2 \approx 0.66$) is provided by a column stochastic matrix of the number of trips $\Lambda_{ij} = \frac{w_{ij}^{\text{trips}}}{\sum_k w_{kj}^{\text{trips}}}$. The column stochastic normalization implies that the total rate at which locations are recruited $\sum_i \Lambda_{ij} = 1$ is constant. The rate of spreading across locations $\beta \approx 3$ is about one and a half times larger than the rate within a location $\gamma \approx 2$. This is somewhat surprising, as we could expect diffusion in the own neighbourhood to be more effective, which apparently is not the case. Mobility is more predictive of participation than communication in this model. This suggests face-to-face contacts are more important for the diffusion than communication over the phone.

In conclusion, the effect of the distance on participation may partly be explained by the social network, either through central recruitment or further diffusion. Nonetheless, the direct contact model provides a better fit overall, making it a better model of participation than the diffusion model. Possibly, this is due to the fact that we only work with an aggregated network, rather than the individual network. However, it is difficult to ascertain whether the decay of participation over distance is entirely due to network effects.

2.2.3. Combining distance and network effects

We want to analyse whether the social network can explain all decay of participation over distance, or whether still some direct effect of distance is left. We therefore incorporate distance also in the network model. In particular, the direct contact model overestimates participation at large distances (see Fig. 3(e)). It therefore makes sense to hypothesise that the probability to participate is not constant, but rather depends on the distance also, so that we arrive at

$$\hat{m}_j = q n_j^{\beta} (1 - (1 - p e^{-\gamma d_{ij}})^{\Lambda_{ij}}), \tag{5}$$

using exponential decay. We find that this model does an excellent job ($R^2 \approx 0.71$ using the number of calls), whereas a powerlaw decay actually worsens the fit (Fig. 3(i)–(l)). This suggests that either long-distance calls are less effective at recruiting people, or that distance is an obstacle, preventing people from going to the protest. The scaling with the number of users is now consistently superlinear for both communication and mobility ($\beta \approx 1.1$), suggesting more populous places are more likely to engage in protests. The distance at which the probability is roughly halved, is about $\log(2)/\gamma \approx 9\,\mathrm{km}$ for both the number of calls and number of trips, while for the duration, this half-distance is about 4 km. As an alternative, we also tested an intervening recruitment model, where nearby users would be called before more distant users, but this model did very poorly. This suggests that distance still acts as an obstacle to participation, even when controlled for network effects.

3. Conclusion

Although much is known about protesters, their geographical origins were still unclear. We here find that distance has a major impact on participation, and that most participants have rather local origins. A simple gravity model already fits the data well. A model based on intervening opportunities did considerably worse. This is consistent with distance acting as an obstacle to participation.

Protesters are usually already acquainted with people in the movement prior to the protest [7]. Since social networks also exhibit spatial patterns [14, 15], this could be an explanation of how participation drops over distance. We indeed find that the social network is predictive of participation. Interestingly, face-to-face contact seems more important for further diffusion of participation, whereas telephone communication works better in a model of central recruitment by the organisers. This is consistent with

earlier studies, where going from door to door, talking to people face-to-face, has a larger effect than contacting people over the phone, by e-mail or through the post [44].

However, the social network only explains part of the decay of participation over distance. Even after controlling for the contact intensity between two regions, there is still an effect of distance. In particular, the probability to participate seems to decay exponentially, with a radius of roughly 9 km to halve the probability to participate. This again suggests that distance functions as an obstacle for participation. Alternatively, this could be because long-distance recruitment is less effective.

Hence, if network data on interaction between locations is available, the direct contact model with exponential decay over distance would be recommended (using any interaction measure). That model provides the best fit overall, and provides the best fit for 22 of the 39 protests. If available data is limited to populations at locations and the distances between locations, only two models are possibly suitable: the gravity model and the intervening recruitment model. The latter does particularly badly, and the recommend model is then the gravity model with powerlaw decay.

Our analysis suggests that participation is affected by three main factors: (1) social network; (2) geography; (3) recruitment potential. These three factors relate to each other in a specific way, quite different from a linear (logistic) regression. The models we use should be a step forward from linearity in understanding how people are recruited for protests. Nonetheless, these models are quite rudimentary, and should serve as a starting point for integrating other effects. How these various effects integrate into the model exactly should be explored further. Different phenomena may affect these three main factors in different ways. For example, it is possible that relative deprivation [45] and socio-economic factors play a role in the recruitment potential, but otherwise don't affect the social network or geography. Similarly, some factors may affect the distance impedance, but not affect the recruitment potential directly, such as city limits, transport infrastructure or natural barriers. Resources may be mobilized [46] to contact people more effectively, altering the impact of the network, or to provide transportation, altering the geographic impact. But those resources do not necessarily increase the recruitment potential itself. This has important consequences for the recruitment process. Different choices may have different consequences for participation. Our results may help guide such choices.

Still, we must be careful in interpreting our results. Perhaps we measure more the (unusual) presence of people at a certain location (which might be closely related to mobility) and not protest participation as such. Therefore,

future research should compare protests to other types of events. The current study only uses aggregated data. Although protest participation seems to diffuse over the network, we cannot assess whether it diffuses through individual contacts or whether it is simply a reflection of the overall interaction between two location. Access to the individual network data would make it possible to study this in greater detail. There are other possible models of diffusion, such as cascading models [47] or threshold models [48], but testing them warrants individual data. We could then also analyse the interaction of mobility and communication better.

Using mobile phone data may introduce a bias towards those owning and using a mobile phone. Perhaps only wealthier people own phones, who tend to live more in urban areas, so that we underestimate participation from rural areas. Similarly, antenna density may bias the results as rural areas have larger distances between nearby antennas. People attending a nearby event may thus be estimated to live further away in rural areas than in urban areas. Moreover, we will miss people that have not communicated during the protest, or that used a different nearby antenna for communication. Similarly, we may possibly identify some people as participants while they are merely spectators. Finally, protests are not necessarily confined to a single location, and more complex mass movements are not captured by our current approach. It would therefore be good if these spatial patterns would be corroborated through other approaches such as social media [31, 32] and traditional approaches (e.g. surveys).

Acknowledgements

VAT acknowledges support by the Royal Netherlands Academy of Arts and Sciences (KNAW) through its eHumanities project. PMAS acknowledges partially support by the Russian Scientific Foundation, grant #14-21-00137. PMAS and RQ acknowledge the financial support of the Future and Emerging Technologies (FET) programme within Seventh Framework Programme (FP7) for Research of the European Commission, under the FET-Proactive grant agreement TOPDRIM, number FP7-ICT-318121, as well as under the FET-Proactive grant agreement Sophocles, number FP7-ICT-317534.

References

[1] B. Klandermans, The Social Psychology of Protest, 1st Edition, Wiley-Blackwell, Oxford, UK ; Cambridge, Mass, 1997.

[2] C. Tilly, The Politics of Collective Violence, Cambridge University Press, Cambridge ; New York, 2003.

[3] C. Tilly, S. Tarrow, Contentious Politics, Oxford University Press, USA, Oxford, Oxon, UK, 2006.

[4] D. S. Meyer, The Politics of Protest: Social Movements in America, 2nd Edition, Oxford University Press, New York, 2014.

[5] R. K. Givan, K. M. Roberts, S. A. Soule, The Diffusion of Social Movements: Actors, Mechanisms, and Political Effects, Cambridge University Press, Cambridge, 2010.

[6] S. A. Soule, Diffusion Processes within and across Movements, in: D. A. Snow, S. A. Soule, H. Kriesi (Eds.), The Blackwell companion to social movements, Blackwell, Oxford, 2004, pp. 294–310.

[7] M. Diani, Networks and Participation, in: D. A. Snow, S. A. Soule, H. Kriesi (Eds.), The Blackwell Companion to social movements, Blackwell, Oxford, 2004, pp. 339–359.

[8] B. Klandermans, D. Oegema, Potentials, networks, motivations, and barriers: Steps towards participation in social movements, Am. Sociol. Rev. 52 (4) (1987) 519–531.

[9] A. Schussman, S. A. Soule, Process and Protest: Accounting for Individual Protest Participation, Soc. Forces 84 (2) (2005) 1083–1108. doi:10.1353/sof.2006.0034.

[10] D. McAdam, R. Paulsen, Specifying the Relationship Between Social Ties and Activism, Am. J. Sociol. 99 (3) (1993) 640–667.

[11] D. McAdam, Recruitment to High-Risk Activism: The Case of Freedom Summer, Am. J. Sociol. 92 (1) (1986) 64–90.

[12] M. Diani, Networks and Social Movements, in: The Wiley-Blackwell Encyclopedia of Social and Political Movements, Blackwell Publishing Ltd, Malden, USA, 2013.

[13] D. A. Snow, L. A. Zurcher Jr, S. Ekland-Olson, Social networks and social movements: A microstructural approach to differential recruitment, Am. Sociol. Rev. 45 (5) (1980) 787–801.

[14] R. Lambiotte, V. D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, P. Van Dooren, Geographical dispersal of mobile communication networks, Physica A 387 (21) (2008) 5317–5325. doi:10.1016/j.physa.2008.05.014.

[15] G. Krings, F. Calabrese, C. Ratti, V. D. Blondel, Urban gravity: a model for inter-city telecommunication flows, J. Stat. Mech. 2009 (07) (2009) L07003. doi:10.1088/1742-5468/2009/07/L07003.

[16] F. Calabrese, F. C. Pereira, G. D. Lorenzo, L. Liu, C. Ratti, The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events, in: P. Floréen, A. Krüger, M. Spasojevic (Eds.), Pervasive Computing, no. 6030 in Lecture Notes in Computer Science, Springer Berlin Heidelberg, Heidelberg, 2010, pp. 22–37.

[17] M. Wirz, T. Franke, D. Roggen, E. Mitleton-Kelly, P. Lukowicz, G. Tröster, Probing crowd density through smartphones in city-scale mass gatherings, EPJ D.S. 2 (1) (2013) 5. doi:10.1140/epjds17.

[18] L. Gao, C. Song, Z. Gao, A.-L. Barabási, J. P. Bagrow, D. Wang, Quantifying Information Flow During Emergencies, Sci. Rep. 4 (2014) 3997. doi:10.1038/srep03997.

[19] J. P. Bagrow, D. Wang, A.-L. Barabási, Collective response of human populations to large-scale emergencies., PLoS ONE 6 (3) (2011) e17680. doi:10.1371/journal.pone.0017680.

[20] V. A. Traag, A. Browet, F. Calabrese, F. Morlot, Social Event Detection in Massive Mobile Phone Data Using Probabilistic Location Inference, in: Proceedings IEEE SocialCom'2011, IEEE, Boston, 2011, pp. 625–628. doi:10.1109/PASSAT/SocialCom.2011.133.

[21] A. Dobra, N. E. Williams, N. Eagle, Spatiotemporal Detection of Unusual Human Population Behavior Using Mobile Phone Data, PLoS ONE 10 (3) (2015) e0120449. doi:10.1371/journal.pone.0120449.

[22] A. Amini, K. Kung, C. Kang, S. Sobolevsky, C. Ratti, The impact of social segregation on human mobility in developing and industrialized regions, EPJ D.S. 3 (1) (2014) 6. doi:10.1140/epjds31.

[23] S. Phithakkitnukoon, Z. Smoreda, P. Olivier, Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data, PLoS ONE 7 (6) (2012) e39253. doi:10.1371/journal.pone.0039253.

[24] C. Song, Z. Qu, N. Blumm, A.-L. Barabási, Limits of predictability in human mobility., Science 327 (5968) (2010) 1018–21. doi:10.1126/science.1177170.

[25] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, C. Mascolo, A Tale of Many Cities: Universal Patterns in Human Urban Mobility, PLoS ONE 7 (5) (2012) e37027. doi:10.1371/journal.pone.0037027.

[26] F. Simini, M. C. González, A. Maritan, A.-L. Barabási, A universal model for mobility and migration patterns., Nature 484 (7392) (2012) 96–100. doi:10.1038/nature10856.

[27] B. C. Csáji, A. Browet, V. A. Traag, J.-C. Delvenne, E. Huens, P. Van Dooren, Z. Smoreda, V. D. Blondel, Exploring the mobility of mobile phone users, Physica A 392 (6) (2013) 1459–1473. doi:10.1016/j.physa.2012.11.040.

[28] M. C. González, C. A. Hidalgo, A.-L. Barabási, Understanding individual human mobility patterns., Nature 453 (7196) (2008) 779–82. doi:10.1038/nature06958.

[29] P. Bajardi, M. Delfino, A. Panisson, G. Petri, M. Tizzoni, Unveiling patterns of international communities in a global city using mobile phone data, EPJ D.S. 4 (1) (2015) 3. doi:10.1140/epjds/s13688-015-0041-5.

[30] V. D. Blondel, A. Decuyper, G. Krings, A survey of results on mobile phone datasets analysis, EPJ D.S. 4 (1) (2015) 10. doi:10.1140/epjds/s13688-015-0046-0.

[31] F. Botta, H. S. Moat, T. Preis, Quantifying crowd size with mobile phone and Twitter data., Royal Society Open Science 2 (5) (2015) 150162. doi:10.1098/rsos.150162.

[32] M. Alanyali, T. Preis, H. S. Moat, Tracking Protests Using Geotagged Flickr Photographs, PLoS ONE 11 (3) (2016) e0150466. doi:10.1371/journal.pone.0150466.

[33] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, C. Ratti, Geo-located Twitter as proxy for global mobility patterns, Cartography and Geographic Information Science 41 (3) (2014) 260–271. doi:10.1080/15230406.2014.890072.

[34] C. Raleigh, A. Linke, H. Hegre, J. Karlsen, Introducing ACLED: An Armed Conflict Location and Event Dataset Special Data Feature, J. Peace. Res. 47 (5) (2010) 651–660. doi:10.1177/0022343310378914.

[35] A. G. Wilson, Complex Spatial Systems: The Modelling Foundations of Urban and Regional Analysis, Pearson Education, Essex, UK, 2000.

[36] R. A. Cochrane, A possible economic basis for the gravity model, Journal of Transport Economics and Policy (1975) 34–49.

[37] A. G. Wilson, A statistical theory of spatial distribution models, Transportation research 1 (3) (1967) 253–269.

[38] D. E. Sherkat, T. J. Blocker, The political development of sixties' activists: Identifying the influence of class, gender, and socialization on protest participation, Soc. Forces 72 (3) (1994) 821–842.

[39] S. Stouffer, Intervening opportunities: a theory relating mobility and distance, Am. Sociol. Rev. 5 (6) (1940) 845–867.

[40] H. E. Brady, K. L. Schlozman, S. Verba, Prospecting for participants: Rational expectations and the recruitment of political activists, Am. Polit. Sci. Rev. 93 (01) (1999) 153–168.

[41] R. R. Schmitt, D. L. Greene, An alternative derivation of the intervening opportunities model, Geogr. Anal. 10 (1) (1978) 73–77.

[42] M. J. Keeling, P. Rohani, Modeling Infectious Diseases in Humans and Animals, 1st Edition, Princeton University Press, Princeton, 2007.

[43] V. Colizza, R. Pastor-Satorras, A. Vespignani, Reaction-diffusion processes and metapopulation models in heterogeneous networks, Nat. Phys. 3 (4) (2007) 276–282. doi:10.1038/nphys560.

[44] A. S. Gerber, D. P. Green, The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment, Am. Polit. Sci. Rev. 94 (3) (2000) 653–663. doi:10.2307/2585837.

[45] J. N. Gurney, K. J. Tierney, Relative Deprivation and Social Movements: A Critical Look at Twenty Years of Theory and Research, Sociol. Q. 23 (1) (1982) 33–47.

[46] B. Edwards, J. D. McCarthy, Resources and social movement mobilization, in: D. A. Snow, S. A. Soule, H. Kriesi (Eds.), The Blackwell companion to social movements, Blackwell, Oxford, 2004, pp. 116–152.

[47] D. J. Watts, A simple model of global cascades on random networks, Proc. Natl. Acad. Sci. USA 99 (9) (2002) 5766–5771. doi:10.1073/pnas.082090499.

[48] M. Granovetter, Threshold models of collective behavior, Am. J. Sociol. 83 (6) (1978) 1420.

[49] K. P. Burnham, D. R. Anderson, Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Springer, New York, 2013.

[50] A. P. Masucci, J. Serras, A. Johansson, M. Batty, Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows, Phys. Rev. E 88 (2) (2013) 022812. doi:10.1103/PhysRevE.88.022812.

Appendix A. Methods

Appendix A.1. Data

Our data consist of anonymised call detail records (CDRs) obtained from a telecom operator from a single undisclosed African country covering a single year somewhere in the last decade. Each CDR included in the dataset details the user who made the call, the date and time of the call and what antenna was used for the call. However, we do not know the recipient of the call, nor what receiving antenna was used. The approximate location of the antenna is provided to us, thus supplying the necessary geographic information. In addition, we have information on an aggregate level for pairs of antennas. For each pair of antennas, we know the number of calls and the total duration of the calls between them. Based on the CDR level data we estimate the number of trips between two antennas.

We combine this mobile phone data set with data from the Armed Conflict Location & Event Data (ACLED) project[4] [34]. This project collects information from various media and (NGO) reports and manually codes events. ACLED provides information on the time and location of events. Unfortunately the timing provided by ACLED is only accurate to the day, so that we cannot use more precise information for determining the presence of users. Similarly, the spatial resolution of the ACLED data depends on

[4] http://www.acleddata.com/

the information available from the source. ACLED does provide as accurate localisations as possible, usually at the level of a (part of a) city. They report the accuracy of this coding and provide exact latitude and longitude coordinates for each event. All 39 protest events studied here have been geolocated at the highest precision level. Nonetheless, this limited spatial and temporal resolution may distort some of the results, and more accurate results may be expected with a more precise coding. Because of the sensitivity of the topic, we do not disclose any information pertaining to the location or timing of the protests to prevent any possible misuse of our results.

Appendix A.2. Mobile phone data processing

The most frequently used antenna indicated the origin of the participants. The population, i.e. the number of users, at a certain location is estimated by looking at the most frequently accessed antenna by a user, which we denote by a_u. We then simply count how many users have a particular antenna as their most frequent antenna. This provides the number of users at antenna i as

$$n_i = |\{u \mid a_u = i\}|. \tag{A.1}$$

We estimate whether somebody is a participant by looking whether (s)he was at the particular location at a particular date, but not often on any other date (taking into account the difference between weekends and weekdays). More formally, let us denote by $c_{uj}(t)$ the number of calls user u had during date t at location j for some protest. We similarly calculate $\mu_{uj} = \langle c_{uj} \rangle$ the average number of calls (where the average is taken over time) and the standard deviation $\sigma_{uj} = \sqrt{\langle c_{uj}^2 \rangle - \langle c_{uj} \rangle^2}$. For each user we calculate a z-score as $z_{uj} = \frac{c_{uj}(t) - \mu_{uj}}{\sigma_{uj}}$. We assume that user u participated to the event if z is higher than a critical threshold z_c, and we call such a user a participant. For the results in the main text we used $z_c = 1.96$, which corresponds to a 0.0025 one-tailed probability in a standard Gaussian distribution. The number of participants coming from antenna i is calculated as the number of users that have antenna i as their most frequent antenna, while they also had a higher z-score than the cut-off:

$$m_i = |\{u \mid a_u = i, z_{uj} > z_c\}|. \tag{A.2}$$

We also tested $z_c = 1.64$ and $z_c = 2.33$ which corresponds to a 0.05 and 0.001 one-tailed probability. In general, a higher cut-off leads to slightly lower fits overall, but the results remain qualitatively the same. To give some idea,

for the largest protest, we detected respectively 1148, 1105 and 1010 for a cut-off of $z_c = 1.64$, $z_c = 1.96$ and $z_c = 2.33$. The average participation decreased from about $\langle m_j \rangle \approx 356$ for $z_c = 1.64$ to $\langle m_j \rangle \approx 340$ for $z_c = 1.96$ and then to $\langle m_j \rangle \approx 314$ for $z_c = 2.33$. Overall, the results are quite robust with respect to this parameter.

We do the above calculations for the number of participants and the number of users for each event separately. In addition, we calculate the number of trips between location i and j based on the observed calls. Whenever we first observe a mobile phone in location i and then in location j we record that as a trip from location i to j. This only makes sense for movements between different antennas, so that the rate of internal movement remains unknown. We count the total number of trips over the whole period for each pair of antennas.

Appendix A.3. Model fitting

Although we report R^2 values to provide some idea of fit, it is a flawed measure, in particular for model selection. We therefore corroborate our findings using the root mean square error (RMSE) and Akaike information criterion (AIC) which is based on information theoretical principles [49]. We use non-linear least-squares to estimate the model parameters, which assumes normally distributed errors $\epsilon_j = m_j - \hat{m}_j \sim \mathcal{N}(0, \sigma^2)$ with zero mean and variance σ^2. We minimize the sum of squared errors $\sum_j \epsilon_j^2$ using Levenberg-Marquardt, but if this does not converge we resort to Nelder-Mead. The RMSE is defined as

$$\sigma = \sqrt{\langle \epsilon^2 \rangle} = \sqrt{\frac{1}{r} \sum_j \epsilon_j^2} \tag{A.3}$$

where r is the number of locations. We can define AIC as

$$\mathrm{AIC} = r \log \sigma^2 + 2(k+1) \tag{A.4}$$

where k is the number of parameters in the model [49]. Better models provide lower RMSE and lower AIC. For assessing the overall performance of a model we calculated the AIC over all protests taken together. Since the AIC has no absolute meaning, we only provide relative differences $\Delta\,\mathrm{AIC}_q = \mathrm{AIC}_q - \min_s \mathrm{AIC}_s$ for model q. It is customary to also provide AIC weights, which are proportional to $\exp -\Delta\,\mathrm{AIC}_q\,/2$ (normalised to sum to 1), but $\Delta\,\mathrm{AIC}_q$ is too large for the total AIC. We do provide the average AIC weights over all protests individually.

All calculations were done in `python` 2.7 with `numpy`, `scipy`, `pandas` and `igraph` packages. The preparation of the data is done using a combination of `gawk`, `sort` and `python`. Plotting was done using `matplotlib` and PGFPLOTS.

Appendix A.3.1. Models

Gravity model The gravity model commonly also uses various distance decay functions. In our case, the more general formulation is

$$\hat{m}_j = Z \frac{n_j^\beta}{f(d_{ij})}, \tag{A.5}$$

where f is one of the following distance decay functions

$$f(d_{ij}) = \begin{cases} e^{\theta d_{ij}} & \text{Exponential} \\ (d_{ij}+\delta)^\gamma & \text{Powerlaw} \\ (d_{ij}+\delta)^\gamma e^{\theta d_{ij}} & \text{Powerlaw w cut-off} \end{cases}. \tag{A.6}$$

Intervening recruiment For the intervening recruitment model in equation (2) we used an exponential version. We did this to increase the numerical stability to estimate the probability $\phi = -\log(1-p)$. This leads to the (otherwise equivalent) formulation

$$\hat{m}_j = Z(e^{-\phi(n_i+n_{ij})} - e^{-\phi(n_i+n_j+n_{ij})}). \tag{A.7}$$

We also included two simplistic models that may serve as a baseline comparison. The first model simply assumes every location is equally likely to participate in the protest, so that $m_i = Z n_i^\beta$, which we call the uniform model. The second model only takes into account the distance so that $m_i = Z/f(d_{ij})$ with a powerlaw decay $f(d_{ij}) = (d_{ij}+\delta)^\beta$, which we term the distance model.

Direct contact For the distance decay in the direct contact model we also use Eq. A.6.

SI model The dynamical system of epidemic spreading needs to be solved numerically. We fix the initial condition at $t = 0$ to a single protester in the protest location and the time of the protest at $t = 1$. The fixing of the time is without loss of generality, since any other t would simply imply a rescaling of the parameters β and γ.

Appendix A.3.2. Model of the network

We also modelled the number of calls w_{ij}^{calls}, the duration of the calls w_{ij}^{duration} and the number of trips between two location w_{ij}^{trips}. The overview of the fitting to the number of calls w_{ij}^{calls}, the duration w_{ij}^{duration} and the number of trips w_{ij}^{trips} is provided in Table A.3.

We test both the traditional gravity model and the recently proposed radiation model. The usual form of the gravity model is

$$\hat{w}_{ij} = Z \frac{n_i^{\alpha} n_j^{\beta}}{f(d_{ij})} \tag{A.8}$$

where $f(d_{ij})$ is a distance decay function as in equation (A.6). The radiation model [26] is specified as

$$\hat{w}_{ij} = Z w_i \frac{n_i n_j}{(n_i + n_{ij})(n_i + n_j + n_{ij})}, \tag{A.9}$$

where $w_i = \sum_j w_{ij}$ is the total out-strength of node i.

We also tested an alternative specification where there would be a home advantage, which consists of artificially increasing the number of people at home, so that

$$\hat{w}_{ij} = Z w_i \frac{(n_i + h) n_j}{(n_i + h + n_{ij})(n_i + h + n_j + n_{ij})}. \tag{A.10}$$

Although this model improved the fit of the radiation model somewhat, it still does worse than the gravity model. This clearly favours the gravity model over the radiation model, as was also found by others [50].

	Δ AIC (Total)	AIC weight (mean)	# best AIC	R^2	RMSE	RMSE improv. (%)
Baseline						
Uniform	113 901	0.00		0.05	1.05	0.0
Distance	57 682	0.00		0.50	0.69	29.3
Gravity Model						
Gravitational Model (exp)	22 625	0.00		0.61	0.55	39.4
Gravitational Model (pl)	16 360	0.03	1	0.63	0.53	41.2
Gravitational Model (pl-exp)	18 025	0.03	1	0.59	0.54	38.6
Intervening Opportunities						
Radiation Model	83 729	0.00		0.31	0.84	16.5
Event Radiation Model	31 442	0.00		0.55	0.60	35.0
Direct Contact						
No distance decay						
Number of calls	6 277	0.02	1	0.68	0.49	45.7
Total Duration	33 266	0.02	1	0.58	0.59	36.8
Number of trips	13 908	0.02	1	0.67	0.51	45.1
Exponential distance decay						
Number of calls	0	0.17	7	0.71	0.46	48.7
Total Duration	10 772	0.30	11	0.70	0.49	47.4
Number of trips	14 534	0.11	4	0.65	0.51	44.4
Powerlaw distance decay						
Number of calls	66 343	0.09	3	0.51	0.67	33.1
Total Duration	86 674	0.05	2	0.44	0.81	25.5
Number of trips	64 929	0.05	2	0.36	0.69	28.9
Epidemic Model						
Number of calls						
Normal	57 903	0.00		0.44	0.71	25.4
Row stochastic	55 006	0.00		0.52	0.68	30.3
Column stochastic	60 079	0.00		0.48	0.70	28.3
Total duration						
Normal	74 262	0.00		0.37	0.79	20.2
Row stochastic	75 002	0.00		0.42	0.78	23.0
Column stochastic	78 102	0.00		0.39	0.80	21.2
Number of trips						
Normal	24 497	0.00		0.59	0.56	38.3
Row stochastic	20 118	0.00		0.64	0.53	42.1
Column stochastic	7 872	0.13	5	0.66	0.50	44.6

Table A.1: **Model fit.** Fitting results for all specification of all models. The Δ AIC contains the difference of the minimal total AIC. A lower Δ AIC thus indicates a better fit. The AIC weight contains the average AIC weight over all protests. That is, for each protest we calculate the weight $\exp -\Delta \mathrm{AIC}_q /2$, normalised to sum to 1, and we then average over all protests. The number of protests that are best fit by a specific model are denoted in the column # best AIC. The RMSE improvement is calculated with respect to the worst model, which is the uniform baseline model.

Baseline	Z	β	γ	δ	
Uniform	$3.4 \times 10^{-3} \pm 2.3 \times 10^{-3}$	1.1 ± 0.20			
Distance	$4.9 \times 10^{3} \pm 1.6 \times 10^{1}$		2.4 ± 0.22	3.9 ± 0.43	
Gravity Model	Z	β	γ	δ	θ
Exponential	1.8 ± 0.53	0.85 ± 0.044	0.74 ± 0.023		
Powerlaw	$6.5 \times 10^{1} \pm 1.8$	0.78 ± 0.044	1.8 ± 0.13	2.7 ± 0.34	
Powerlaw with cutoff	0.29 ± 0.12	0.77 ± 0.046	1.7 ± 0.10	0.83 ± 0.097	$0.21 \pm 5.7 \times 10^{-3}$
Intervening Opportunities	Z	ϕ			
Radiation Model	$3.4 \times 10^{1} \pm 0.99$				
Event Radiation Model	$2.0 \times 10^{2} \pm 6.1$	$6.8 \times 10^{-4} \pm 2.2 \times 10^{-5}$			
Direct Contact					
No distance decay	q	β	p		
Number of calls	0.076 ± 0.015	1.2 ± 0.042	$0.010 \pm 1.3 \times 10^{-3}$		
Total duration	0.071 ± 0.021	1.2 ± 0.044	$0.017 \pm 1.7 \times 10^{-5}$		
Number of trips	0.17 ± 0.039	0.91 ± 0.040	$0.19 \pm 7.2 \times 10^{-3}$		
Exponential distance decay	q	β	p	γ	
Number of calls	0.21 ± 0.072	1.1 ± 0.043	0.042 ± 0.012	$0.079 \pm 8.7 \times 10^{-3}$	
Total duration	0.18 ± 0.069	1.2 ± 0.044	$1.7 \times 10^{-3} \pm 6.9 \times 10^{-4}$	0.15 ± 0.012	
Number of trips	0.57 ± 0.20	1.1 ± 0.048	0.085 ± 0.013	$0.071 \pm 8.2 \times 10^{-3}$	
Powerlaw distance decay	q	β	p	γ	δ
Number of calls	$0.026 \pm 7.4 \times 10^{-3}$	1.1 ± 0.080	0.043 ± 0.014	0.20 ± 0.13	1.3 ± 0.83
Total duration	0.047 ± 0.032	1.1 ± 0.10	$0.039 \pm 6.7 \times 10^{-5}$	0.35 ± 0.091	$1.6 \times 10^{3} \pm 0$
Number of trips	0.099 ± 0.013	1.0 ± 0.068	0.068 ± 0.024	-0.16 ± 0.16	4.4 ± 1.2
Epidemic Model	β	γ			
Number of calls					
Normal	$4.5 \times 10^{-3} \pm 1.3 \times 10^{-4}$	2.6 ± 0.068			
Column stochastic	2.3 ± 0.069	2.5 ± 0.062			
Row stochastic	3.1 ± 0.076	2.5 ± 0.064			
Total duration					
Normal	$7.1 \times 10^{-5} \pm 2.4 \times 10^{-6}$	2.6 ± 0.071			
Column stochastic	2.1 ± 0.071	2.7 ± 0.065			
Row stochastic	2.9 ± 0.082	2.6 ± 0.069			
Number of trips					
Normal	$0.052 \pm 1.4 \times 10^{-3}$	2.1 ± 0.071			
Column stochastic	2.8 ± 0.13	1.4 ± 0.12			
Row stochastic	3.0 ± 0.077	2.0 ± 0.070			

Table A.2: **Estimated coefficients.** Average of the estimates of coefficients for all specifications of the models, including standard errors. We excluded outlier estimates for which the standard error exceeded the coefficient, because they heavily skewed the average.

	Number of calls		Duration		Number of trips	
	Δ AIC	RMSE	Δ AIC	RMSE	Δ AIC	RMSE
Gravity Model						
Exponential	188 764	12 600	300 570	559 647	38 451	27 786
Powerlaw	3 060	11 999	10 123	518 517	467	26 464
Powerlaw with cutoff	0	11 990	0	517 139	0	26 448
Radiation Model						
Original	584 081	13 979	630 893	610 404	129 471	31 231
Home Advantage	502 585	13 683	453 136	582 544	114 880	30 651

Table A.3: **Network fit.** Fit of models to network variables. The gravity model with powerlaw (with cutoff) decay is clearly favoured for all variables. The cutoff goes in the wrong direction however (leading to an exponential increase rather than decrease), so that the powerlaw remains preferable.

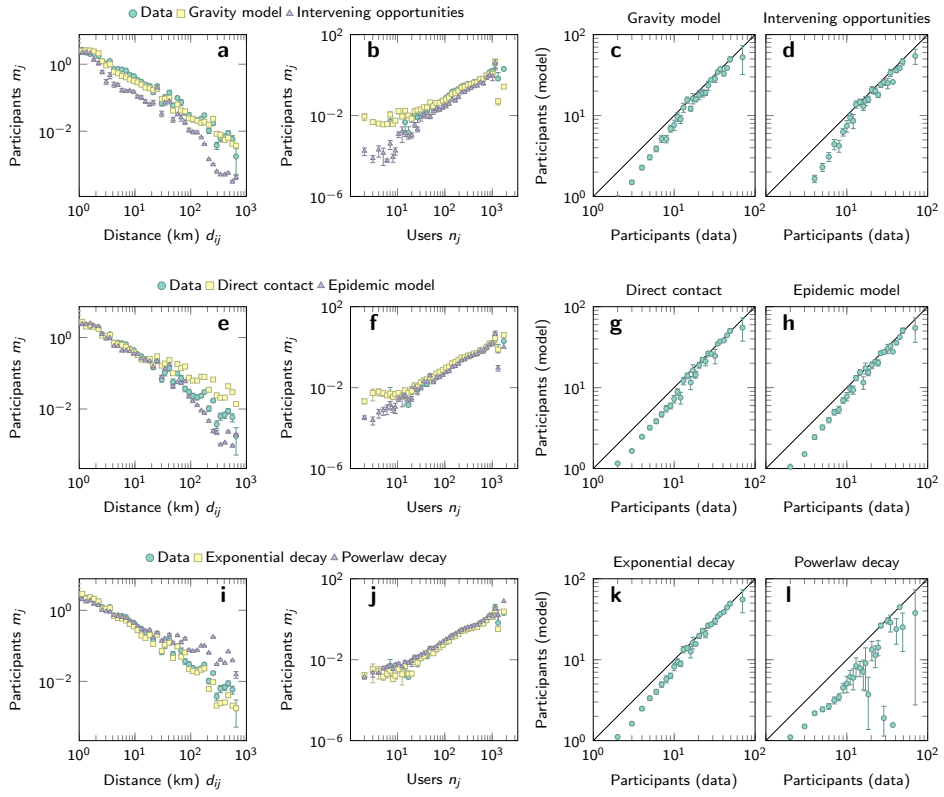


Figure 3: **Model fit.** The fit of the geographical models is displayed in the first row, (a)–(d), the fit of the network models in the second row, (e)–(h), and the fit of the direct contact model with distance dependence in the third row, (i)–(l). In the first column, (a), (e) and (i), we show how the number of participants m_j depend on the distance d_{ij} and in the second column (b), (f) and (j) we show the dependence on the number of users n_j. How well the number of participants is predicted by the various models is displayed in the third and fourth column, (c), (g), and (k) and (d), (h) and (l) respectively. Error bars indicate standard errors of the mean. Clearly, protest participation is affected by distance, and this is well fit by a gravity model ($R^2 \approx 0.63$). The intervening opportunities theory does reasonably well ($R^2 \approx 0.55$) but less well than the gravity model. The geographical dependency can be partly explained by the social network. Assuming the organisers recruit from the protest location and contact people directly over the phone we arrive at the direct contact model, which provides a good fit to the data ($R^2 \approx 0.68$). Assuming that the probability in the direct contact model depends on distance, we find that an exponential decay does best ($R^2 \approx 0.71$) whereas a powerlaw decay does quite poorly ($R^2 \approx 0.51$), both using the number of calls. If participants themselves also recruit people, we obtain an epidemic model which predicts participation also quite well ($R^2 \approx 0.66$). However, whereas phone calls perform best in the direct contact model, mobility is better in the epidemic model. This suggests that face-to-face contact may be more important for spreading participation.

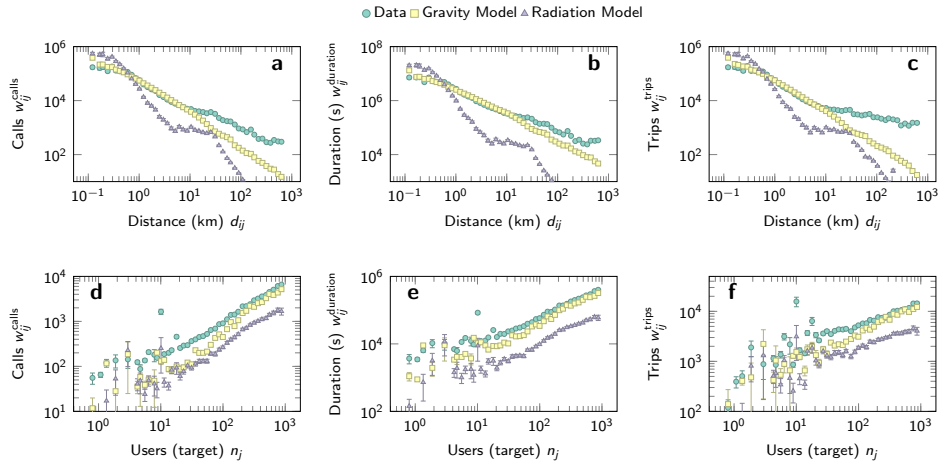


Figure 4: **Social network fit.** Social networks generally depend on distance as well. We show the distance dependency in the first row, (a)–(c), and the dependence on the number of users in the second row, (d)–(f), for respectively the number of calls w_{ij}^{calls} (first column), the duration of the calls w_{ij}^{duration} (second column) and the trips w_{ij}^{trips} (third column). The gravity model provides quite a good fit ($R^2 \approx 0.67$, $R^2 \approx 0.60$ and $R^2 \approx 0.66$ respectively). They all roughly scale as $w_{ij} \sim \frac{\sqrt{n_i n_j}}{(d_{ij}+\delta)^\gamma}$, with $\gamma \approx 1.18$ for the number of calls and the number of trips, while the duration decays much slower with $\gamma \approx 0.86$. Since social networks are important for recruiting people, it is possible that the origins of protesters do not depend on distance as such, but can rather be explained by the social network.