Diffusion of treatment in social networks and mass drug administration

Goylette F. Chami1,2, Andreas A. Kontoleon1, Erwin Bulte1,3, Alan Fenwick4, Narcis B. Kabatereine4,5, Edridah M. Tukahebwa5 & David W. Dunne2

Information, behaviors, and technologies spread when people interact. Understanding these interactions is critical for achieving the greatest diffusion of public interventions. Yet, little is known about the performance of starting points (seed nodes) for diffusion. We track routine mass drug administration—the large-scale distribution of deworming drugs—in Uganda. We observe friendship networks, socioeconomic factors, and treatment delivery outcomes for 16,357 individuals in 3491 households of 17 rural villages. Each village has two community medicine distributors (CMDs), who are the seed nodes and responsible for administering treatments. Here, we show that CMDs with tightly knit (clustered) friendship connections achieve the greatest reach and speed of treatment coverage. Importantly, we demonstrate that clustering predicts diffusion through social networks when spreading relies on contact with seed nodes while centrality is unrelated to diffusion. Clustering should be considered when selecting seed nodes for large-scale treatment campaigns.
The social connections between people influence the spread of information, behaviors, and technologies. Understanding the structure of social networks is essential for targeting beneficiaries and improving the diffusion of public interventions. A key policy challenge concerns how best to identify and to evaluate the performance of starting points (seed nodes) for interventions. Seed nodes in social networks are predominantly selected and targeted based on their centrality. Centrality is a set of measurements that describes the connectivity and embeddedness of a node in a network and is widely viewed as a measure of power and influence in social systems. High centrality of seed nodes promotes simple epidemic spreading, which is the most common model of diffusion. In these models, each additional social contact increases the probability of information transmission and diffusion unfolds along the shortest paths in a network. Long, weak links that bridge otherwise disconnected groups facilitate diffusion, but this simple epidemic spreading does not account for possible brokerage properties of bridges in social networks. On the other hand, the reach of seed nodes has been shown to rely on the topology of the network. Influential seed nodes may group together or belong to the core of the network. Network topology is particularly important in models of complex contagion, where repeated exposure and social reinforcement are needed before an individual partakes in an intervention. In contrast to simple epidemic models, the probability of information or technologies spreading is zero until an exposure threshold is met for the recipient and diffusion tends to travel along short densely clustered ties.

Clustering is one indicator of network transitivity and is pervasive in social networks. Transitivity implies that two nodes are connected if they share a network neighbor. In friendship networks, clustering exists if two friends of the node of interest also are friends. Importantly, in simple epidemic models, high network transitivity causes redundancy and slows spreading, whereas in complex contagion models, this redundancy is required for diffusion to occur.

It is not known whether centrality (simple epidemic) or transitivity (complex contagion) is relevant for evaluating seed node performance in real-world contexts that require repeated contact with seed nodes. Diffusion that relies on repeated contact with seed nodes applies to a number of important social interventions, which is the most common model of diffusion. In these models, each additional social contact increases the probability of information transmission and diffusion unfolds along the shortest paths in a network. Long, weak links that bridge otherwise disconnected groups facilitate diffusion, but this simple epidemic spreading does not account for possible brokerage properties of bridges in social networks. On the other hand, the reach of seed nodes has been shown to rely on the topology of the network. Influential seed nodes may group together or belong to the core of the network. Network topology is particularly important in models of complex contagion, where repeated exposure and social reinforcement are needed before an individual partakes in an intervention. In contrast to simple epidemic models, the probability of information or technologies spreading is zero until an exposure threshold is met for the recipient and diffusion tends to travel along short densely clustered ties.

Clustering is one indicator of network transitivity and is pervasive in social networks. Transitivity implies that two nodes are connected if they share a network neighbor. In friendship networks, clustering exists if two friends of the node of interest also are friends. Importantly, in simple epidemic models, high network transitivity causes redundancy and slows spreading, whereas in complex contagion models, this redundancy is required for diffusion to occur.

M.D.A. is the community-wide distribution of preventive chemotherapies to treat human parasitic worms. Treatments are provided at no financial cost and are safely administered to infected and uninfected individuals. Annual or biannual M.D.A. is necessary due to rapid reinfection. In sub-Saharan Africa, over 633 million people require preventive chemotherapy, yet only 1/3rd of these individuals are approached for treatment through M.D.A. Each village in our study area in Mayuge District, Uganda has two community medicine distributors (C.M.D.s). These individuals are unpaid volunteers, who receive only a nominal reimbursement for transport to training. C.M.D.s are trained and retrained annually to treat all eligible individuals in their village. In 2013, an undisturbed, routine round of M.D.A. was tracked for 1 month. This time frame enabled comparisons across study villages, minimized recall bias for treatment coverage outcomes, and was within national treatment schedules. Data were collected in 17 villages for 16,357 individuals within 3491 homes. Friendship networks were defined at the household level because C.M.D.s were instructed by the local government to move from home to home. An undirected edge between two households was generated if a member of one household named any member of another household as a close friend. The advantages of studying close friendship networks for M.D.A. include the lack of predefined roles and network positions of C.M.D.s, as well as an approximation of static networks in our rural villages. A static network has edges that do not change frequently (in this case, close friendships remained over several years).

Our study design avoids several potential sources of bias and provides a real-world environment to examine seed node performance. M.D.A. for intestinal schistosomiasis and soil-transmitted helminths began in our study area in 2003. Accordingly, the target population was familiar with M.D.A., social learning was unnecessary, and most individuals would comply with treatment if offered in this context. Before the first round of M.D.A. in our study area, government health workers visited communities and asked villagers to meet and select C.M.D.s. This context is the real setting that public interventions must be assessed. Selecting any person, e.g., via random assignment, to be a seed node often is not possible. To integrate stakeholders of interventions, government or non-governmental organizations utilize individuals selected by their communities. Researchers and village members knew the starting points for information and treatment diffusion. The current C.M.D.s had been active for several years and were reminded of their duties during annual retraining. C.M.D.s were well acquainted with their responsibilities and identifiable by other villagers. Apart from C.M.D.s, eligible individuals had no access to M.D.A. pills or information about drug availability outside of their village. Nationally led information campaigns and school-based distribution had not commenced during the study. Hence, information on drug availability was initially monopolized by C.M.D.s. Per routine M.D.A. procedures, C.M.D.s received training and drugs for M.D.A. outside of their village. Treatments were only available from C.M.D.s within the village. Nearby health centers and primary schools had not received drugs during the study (due to delays by the national control program) and C.M.D.s did not leave pills with treated households to pass onwards to untreated households. Also, C.M.D.s had knowledge of and registered all households prior to the start of distribution and were provided a sufficient number of pills to treat all individuals in their village.

The treatment coverage achieved by C.M.D.s was examined. In each village, we measured coverage as the percentage of eligible households that were visited and offered treatment by C.M.D.s. Specifically, household coverage (binary) was defined as at least one eligible person in the home being approached by C.M.D.s and offered at least one treatment through M.D.A. Positive household coverage included eligible individuals who were visited by C.M.D.s and refused treatment. Treatment coverage measured the delivery of medicines. Coverage was studied to isolate C.M.D. performance (as opposed to the ingestion of pills; compliance) simply because the choice to offer medicine to eligible individuals rests with the C.M.D., whereas compliance with the offered treatment is dependent on the final decision of the recipient. To identify what factors determined the total reach and speed of household coverage in each village, we assessed measures of centrality and local transitivity of seed nodes. These network indicators were...
normalized to account for variation in village/network size and then averaged for the two CMDs in each village. For the speed of diffusion, the number of days required for CMDs to reach 50% household coverage was examined. This 50% target is a highly conservative measure of World Health Organization (WHO) guidelines, which recommend at least 75% coverage of all school-aged children as a feasible goal. The 50% target was chosen to include as many villages as possible while remaining as close as possible to actual WHO goals. We find that clustering predicts the reach and speed of drug delivery, while centrality is unrelated to diffusion. Clustering should be considered when selecting seed nodes for large-scale treatment campaigns.

**Results**

**Network completeness and external consistency.** Complete, sociocentric networks were constructed that included a total of 3436 nodes and 16,155 edges (Supplementary Table 1). In all
villages, if the household did not belong to the main connected component, the household was an isolate (belonged to no other network components). Across the 17 villages, only 55 households were not in the main component and 87 households refused to participate in the survey. Therefore, 96.03% (3436/3578) of all households available at the time of survey were included in the main component of their village network. Only 4.95% (859/17347) of names provided in all villages for friendship connections—the maximum possible edges, including self-loops and multi-edges—were not matched; a remarkable 99.04% (16200/16357) of individuals who were recorded in household surveys were in a household included in a main network component. No upward bias, which would entail individuals naming as many outgoing connections as possible irrespective of friendship status, was observed. Only 4.83% (166/3436) of households included in the friendship networks listed the maximum of 10 outgoing connections.

The networks studied here share structural properties with a wide range of complex physical, biological, and social systems (Supplementary Table 1 and Supplementary Fig. 1)32. High centrality was observed for CMDs (Fig. 1), who were selected by members of their village. The degree of each CMD varied considerably and ranged from 7 to 80. However, CMD degree was greater, in every village, than the average degree of all households. As shown in Fig. 2, wide variation was observed in the local transitivity of connections among CMDs and their friends. The clustering coefficient of CMDs varied greatly from star-shaped ego networks (clustering coeff. = 0; village ID 1) to highly transitive ego networks (clustering coeff. = 0.429; village ID 17). See Supplementary Table 2 for a summary of CMD network statistics.

**Treatment coverage achieved.** In the study villages, household coverage varied from 23.14 to 93.53% (Supplementary Tables 3, 4). Overall, approximately one-third (1011/3415) of households were not approached by CMDs during 1 month of MDA. There was little or no informational diffusion about drug availability between households that were visited by CMDs and households that were never approached for treatment. When non-recipients were asked if they heard that drugs were available...
during the 1-month distribution, 85.06% (860/1011) of households that were not offered medicines by CMDs had no eligible person in the home that knew drugs were available in the village. These households constituted 25.18% (860/3415) of all households with at least one person who was eligible for treatment through MDA.

Clustering, due to its positive correlation with the number of households visited by CMDs, was associated with increased treatment coverage. The frequency that friends of CMDs were also friends (clustering) was positively related to household coverage (\(p\)-value = 0.011) (Table 1). CMDs in the 90th percentile of clustering were estimated to have approached 9.32% (\(p\)-value < 0.001) more households in their village, achieving on average 78.1% household coverage, when compared to CMDs in the 10th percentile of clustering who reached an estimated 68.8% of households. Hence, clustering was associated with a village realizing the 75% coverage target set by the WHO\(^25\). This target is an approximation of the treatment coverage needed to control morbidity attributable to helminthic infections. Our results are consistent with findings from artificially constructed online networks. Centola shows how global (network-level) clustering is efficient for the spread of social behaviors\(^1\). Global (village level) clustering was associated with average CMD clustering (Supplementary Fig. 2) and household coverage (Supplementary Table 5). Though, when global and CMD clustering were compared in stepwise regressions, using an alpha-to-enter of 0.15, only average CMD clustering was retained as a significant predictor of household coverage.

Although CMDs with high clustering achieved the greatest household coverage, the number or completeness of connections among their friends was not informative. The proportion of possible edges (density) was borderline insignificant (\(p\) value = 0.055) and the fraction of reciprocated ties among friends of CMDs was uncorrelated (\(p\) value > 0.05) with household coverage.

Belonging to a well-connected group of friends was not associated with household coverage; the largest degree shared among the friends of CMDs (core number) was insignificant (\(p\) value > 0.05). There was no support (\(p\) value > 0.05) that CMDs needed to belong to a large set of households, where everyone was completely connected to everyone else in the friend group, i.e., have a high clique number.

Social network analysis in low-income countries has relied on centrality for guiding the selection of seed nodes\(^8\). For example, in a project in India, seed nodes were selected non-randomly by a non-governmental organization and used to spread information about an innovation—microfinance\(^7\). The communication and eigenvector centralities of these first-informed seed nodes were positively associated with the reach of diffusion, although which seed nodes actually shared information was unknown. We found no such relationship in our study. CMD centrality indicators, which measured neighborhood size, connectivity, embeddedness, and accessibility to other nodes, were unrelated (\(p\) value > 0.05) to household coverage. Our results suggest that being well connected or embedded in a network may facilitate the diffusion of innovations with simple epidemic spreading\(^7\), but was not informative for the diffusion of a familiar technology, where spreading relied on contact with seed nodes (our study). There is experimental support for this inference from Honduras, where well-connected nodes did not achieve greater diffusion of multivitamins when compared to randomly selected nodes\(^3\). We also find no correlation of clustering with centrality (Supplementary Table 6). Clustering as opposed to centrality should be considered when selecting seed nodes for existing technologies and diffusion processes that depend on repeated contact with seed nodes.

**Speed of treatment distribution.** The speed of treatment distribution was assessed by comparing the number of days

---

**Table 1 Influence of CMD network properties on village treatment coverage**

| Average CMD\(^5\) network properties | (A) Dependent variable: household coverage\(^b\) | (B) Estimated difference for CMDs in 10th and 90th %iles of clustering\(^a\) |
|---------------------------------------|---------------------------------------------|---------------------------------------------|
|                                       | Coeff. | Robust std. error | \(p\) value | 95% CI          | Coeff. | Robust std. error | \(p\) value | 95% CI          |
| Clustering                           | 2.107  | 0.829             | 0.011       | 0.482          | 3.733  |
| Reciprocity of CMD ego-network\(^d\) | 53.521 | 34.511            | 0.121       | -14.119        | 121.162|
| Density of CMD ego network           | 4.175  | 2.172             | 0.055       | -0.082         | 8.431  |
| Degree                               | -0.008 | 0.009             | 0.362       | -0.024         | 0.009  |
| Avg. neighbor degree                 | -0.020 | 0.030             | 0.496       | -0.078         | 0.038  |
| Core number                          | -0.025 | 0.056             | 0.657       | -0.135         | 0.085  |
| Cliques number                       | 0.021  | 0.103             | 0.842       | -0.182         | 0.223  |
| Closeness                            | 0.520  | 1.350             | 0.700       | -2.125         | 3.166  |
| Eigenvector                          | -0.061 | 1.630             | 0.970       | -3.256         | 3.134  |
| Katz                                 | 0.407  | 1.199             | 0.734       | -1.942         | 2.756  |
| Betweenness                          | -2.795 | 4.109             | 0.496       | -10.848        | 5.259  |
| Communicability                      | -0.139 | 0.571             | 0.808       | -1.258         | 0.980  |

\(^a\)Each model is a single predictor regression. Fractional response models were used with probit links; constants are not presented.

\(^b\)Household coverage is defined as the proportion of households in the village where at least one eligible person in the home was offered at least one treatment through mass drug administration.

\(^c\)CMD is an abbreviation for community medicine distributor. The averages for the two CMDs in each village were used as predictors in the univariate regressions.

\(^d\)Adjusted for heteroskedasticity.

\(^e\)Panel B presents the estimated margins from the model presented in panel A. An increase in clustering for CMDs in the 10th percentile to the 90th percentile would improve household coverage by 9.32%.
required for each village to achieve 50% household coverage (Fig. 3). Two of five villages (ID 7 and ID 17) that achieved the highest household coverage were also among the five fastest villages. The frequencies of households offered treatment against the day of CMD visits are provided in Supplementary Table 7. Three indicators of local transitivity were associated with the speed of treatment distribution—clustering, reciprocity, and density (Table 2). CMDs with tightly knit friends (high clustering) were more likely ($p = 0.015$) to approach 50% of households sooner than CMDs with friends who were not connected (low clustering). The fraction of reciprocated edges in the CMD ego networks was associated ($p = 0.045$) with faster diffusion of household coverage. CMDs with higher edge density in their ego networks were more likely ($p = 0.013$) to achieve 50% household coverage in fewer days than CMDs whose friends were less densely connected. These relationships imply that increasing local transitivity for CMDs may halve the time needed to reach 50% household coverage in a 1-month period of MDA. The predicted time to reach 50% household coverage was 10.48–11.93 days shorter ($p < 0.001$) for CMDs at the 90th percentiles of average clustering, reciprocity, or density when compared to CMDs at the 10th percentiles of these network properties. These results suggest that seed nodes are most efficient, with respect to diffusion speed, when located within clusters. This practical finding can help policymakers conceptualize the time needed to reach targets when diffusion is reliant on clustering. We find that the delivery of medicines occurs much slower than recommended by WHO guidelines, which suggests that treatment coverage of 75% of individuals is a feasible goal in routine schedules set by local governments for MDA. In our study region in Uganda, these routine schedules aim to complete drug distribution within 2–3 weeks (Supplementary Methods). Yet, during 1 month of MDA, only 64.71% (11/17) and 25% (3/12) of villages had CMDs who approached, respectively, 75% of households or 75% of eligible individuals for at least one treatment.

The core and clique numbers of CMDs as well as several indicators of CMD centrality were uncorrelated ($p > 0.05$) with the number of days required to reach 50% household coverage. All results remained when reanalyzed against the proportion of eligible individuals offered treatment (Supplementary Table 8 and Supplementary Fig. 3).

**Robustness to confounders.** We studied possible confounding factors (Supplementary Table 9). Temporal clustering, which was measured as the standard deviation of the day of drug receipt for households connected to CMDs, was insignificant ($p > 0.05$) for household coverage. Thus, the influence of clustered friends on CMD performance cannot be attributed to CMDs quickly and directly communicating with all of their friends about ongoing MDA. Differences in diffusion among villages were not simply

Fig. 3 Time series of household treatment coverage. In Fig. 3, household coverage is the cumulative fraction of eligible households where at least one eligible individual was offered at least one drug through mass drug administration. Information on the day of drug receipt was unavailable for individuals refusing treatment. These non-compliers were excluded from the calculation of treatment delivery speed (see Supplementary Methods). To enable a comparison with the reach of household coverage, the plots are ordered by household coverage achieved, which is stated in the graph titles. Household coverage may differ from the graph maximum. Household coverage included non-compliers and individuals who were offered and accepted treatment but did not remember the day of drug receipt. In parentheses, the village ID is presented. The dashed line represents 50% household treatment. Village ID 17, which had a shorter distribution span than the other villages, was the smallest village of only 65 households. This village also approached 50% of households in the first day of distribution. As a robustness check, the speed analysis was repeated for individual coverage, which was the proportion of eligible individuals offered treatment against the reach of household coverage (Supplementary Fig. 3 and Supplementary Table 8). In this case, village ID 17 approached 50% of eligible individuals on Day 7.
due to the lack of communication between CMDs. The CMDs, who achieved the greatest diffusion, did not need to be close friends. An edge (direct connection) between seed nodes was insignificantly ($p$ value $>0.05$) associated with household coverage. The maximum shortest path length (geodesic distance) connecting two CMDs was three edges and variation in this small social distance was uncorrelated ($p$ value $>0.05$) with household coverage. Incentives to conform or to receive financial payment were highly unlikely as CMDs were unpaid volunteers who were trained to promote MDA. No apparent difference in the abilities of CMDs to distribute treatment was observed. All CMDs were better off with respect to socioeconomic status and network position than the rest of their village (Supplementary Tables 3, 4) and the capacity of CMDs, as measured by degree or closeness centrality, was uncorrelated with diffusion (Tables 1, 2). Hence, community selection successfully identifies well-connected and better-educated people to be CMDs, but can be improved by also incorporating clustering into the selection criteria.

We also explored whether the presence of homophily—the shared likeness among CMDs and their friends—confounded and inflated effects of CMD network properties. We found no evidence ($p$ value $>0.05$) that homophily affected the reach of household coverage (Supplementary Table 9). All results of temporal clustering, CMD friendship, and homophily remained robust against village treatment distribution speed (Supplementary Table 10). Moreover, the shared likeness and personal characteristics of CMDs in each village also did not affect our results or CMD clustering, respectively (Supplementary Tables 11–13).

Lastly, Fig. 4 and Supplementary Table 14 present the association of village size, accessibility, and ecology with treatment coverage and speed. The total number of households was unrelated ($p$ value $>0.05$) to the household coverage achieved and was borderline insignificant ($p$ value $=0.054$) for how quickly 50% of households were approached. Moreover, the proportion of total households directly connected to CMDs in the village friendship network was uncorrelated ($p$ value $>0.05$) with household coverage and treatment delivery speed. Villages with few households (e.g., IDs 9, 15) were not necessarily small with respect to geographical spread (Supplementary Fig. 4). The physical size of the village, which was measured as the furthest distance in meters between two households (spatial diameter), was unrelated ($p$ value $>0.05$) to diffusion. Concerning village accessibility, no correlation ($p$ value $>0.05$) of household coverage or treatment delivery speed with the average distance in meters between two households or village ecology was observed. In addition to moving from home to home to distribute treatments, as instructed by the national program during training, 41.18% (7/17) of villages had CMDs who also informed individuals to come to their home to pick up treatments. This increased availability or access to treatments was uncorrelated ($p$ value $>0.05$) with diffusion.

**Discussion**

Seeding effective distribution of treatment by CMDs relies on clustering. Strong ties among friends of CMDs were associated with an increased rate and level that households were visited and offered treatment by CMDs. Unlike simple epidemic processes that require only one contact to propagate information, seed nodes required constant reinforcement to engage in effective drug distribution. Social reinforcement to deliver treatment is needed for CMDs, who are volunteers, due to the costly investment of time required to visit and revisit all households. Ex-post focus groups with CMDs and village leaders revealed one mechanism of
Fig. 4 Village size and accessibility against treatment outcomes. Household coverage is defined as the proportion of households in the village where at least one eligible person in the home was offered at least one treatment through mass drug administration. Treatment speed is defined as the number of days until community medicine distributors visited 50% of households. Within the one-month distribution period, 15/17 villages achieved 50% household coverage. Supplementary Table 14 presents the regression results and significance of the relationships shown in Fig. 4.

Methods

Clustering can guide public interventions for MDA. MDA programs, in particular for onchocerciasis, have demonstrated success by utilizing kinship structures in rural communities. The selection of CMDs by kinship has reduced CMD attrition, increased treatment coverage, and held CMDs accountable for providing health information. Additional work should investigate the extent in which such social structures may overlap with friendship networks. Friendship networks are another social structure that can complement kinship and be used to improve CMD selection. During the initial selection of seed nodes, village-nominated CMDs can be surveyed to assess if their friends are connected. The final selection of CMDs can be informed by how tightly knit the connections are among their friends. In villages with established CMDs, interventions to improve treatment coverage can focus on increasing clustering. Joint social activities or shared tasks can be established for friends of CMDs to foster connections within this group. These friends may also be asked to serve as a feedback loop of information to CMDs about community perceptions of ongoing problems with drug delivery. This potential role of friends in monitoring the village may partially alleviate opportunity costs that are incurred by volunteer CMDs from time spent distributing medicines.

Methods

Ethics. This study was reviewed and approved by the Uganda National Council of Science and Technology, the Office of the President in Uganda, and the Cambridge University Human Biological Research Ethics Committee. Informed consent was obtained from all respondents and recorded on tablet computers preceding the household interview. All village names have been replaced with village codes to preclude the identification of individuals in the study villages.

Participants and data source. On 1 November 2013, household surveys commenced in the 17 study villages following 1 month of undisturbed MDA. CMDs...
were not informed of these visits beforehand to avoid influencing distribution and the surprise visits were explained as an effort to understand the efficacy of the national drug programs on infection prevalence. Households in each village were approached for interview and information was collected on every person in the home that was at least 1 year of age. The household head was interviewed and, if married, a spouse also was asked to participate. Socioeconomic information and treatment receipt outcomes were surveyed for every household member, and network data were collected at the household level. Data were collected on a total of 16,357 individuals in 3491 households. Infants were not included in the household list. Preliminary work, in February 2013, to understand naming conventions (how children are named when born, how female names change when they marry, and so on) showed that adults did not count individuals younger than 1 year as a household member. Moreover, individuals younger than 1 year were ineligible for any of the drugs that were distributed through MDA, per WHO guidelines.46

**Fig. 5** Schematic of the statistical analysis. The main outcome of interest was diffusion. This outcome was measured as the maximum fraction of eligible households or individuals offered treatment by community medicine distributors (CMDs) and the speed at which those households or individuals were offered treatment. Speed measured which villages reached a specific level of treatment coverage and at which day in the distribution period. The network characteristics of CMDs were the main predictors of interest. To understand why certain network indicators of CMDs may be associated with the diffusion outcomes, robustness checks were conducted. The contribution of homophily to diffusion was assessed, i.e., the affinity for treatment or shared factors that may predict treatment coverage and are unrelated to network topology. The robustness checks also included examining village ecology, size, and physical spread of households (in meters) against diffusion.

**Network construction.** The nodes and edges in the network, respectively, were households and friendship ties. An edge was generated between households if any person in a household was named or had named any person in another household as a close friend. A self-loop occurred when a household named someone within its own home as a close friend. Households were instructed to name individuals outside of their home. Accordingly, self-loops were ignored and only households belonging to the main component were analyzed. Multiple connections (multi-edges) in the same direction from one household to another household, e.g., naming multiple people from the same household, were measured as one edge. We make the working assumption that villages are independent; though it is possible some individuals may have friends in other villages in the study. The full set of 16,357 individuals in the 17 villages was used in the network construction. The name generator was the following:

*Please tell me the clan name first then the second name of up to 10 people that are very close friends to you. You should feel comfortable to turn to this person to borrow tools for fishing or farming without paying. A close friend is also someone that you see frequently. Do not name anyone in your household. Provide the names in the order of who is your closest friend first. Only name people in your village.*

**Statistical analysis.** Treatment coverage, which represented the performance of CMDs, was analysed using Stata v.13.1. Figure 5 presents the diffusion outcomes for CMDs and an overview of the statistical analysis. Diffusion achieved by CMDs was measured at the village level by examining the fraction of the village that was offered treatment by CMDs and the speed at which treatment was distributed. These variables, as defined in the treatment coverage section of the Supplementary Methods, were constructed for both households and individuals to capture different aspects of diffusion and to test the consistency of the results. Household coverage, which measured if at least one eligible individual in the home was offered treatment by CMDs, indicated if the CMD approached the home. Household coverage was the focus of this paper. As a robustness check, individual coverage also was examined. Individual coverage represented the proportion of eligible individuals offered treatment by CMDs in a village. This variable may capture if a CMD returned to a home to provide treatment to individuals who were not present on the first visit or to administer the second round of treatment (albendazole and ivermectin). Topological characteristics of centrality and clustering for the CMD households in the full village friendship network were used to explain diffusion in robust fractional response models with Probit estimation and Poisson regressions. Additional robustness checks are provided in the Supplementary Methods and Supplementary Tables 16–21.
Data availability. All relevant data are available in the paper, Supplementary Information, and on request from the corresponding authors.

Received: 18 May 2016 Accepted: 21 September 2017 Published online: 05 December 2017

References

1. Centola, D. The spread of behavior in an online social network experiment. Science 329, 1194–1197 (2010).
2. Centola, D. An experimental study of homophily in the adoption of health behavior. Science 334, 1269–1272 (2011).
3. Aral, S. & Walker, D. Identifying influential and susceptible members of social networks. Science 337, 337–341 (2012).
4. Bond, R. M. et al. A 61-million-person experiment in social influence and political mobilization. Nature 489, 295–298 (2012).
5. Rogers, E. M., Diffusion of Innovations 5th edn, Vol. 1 (FreePress, 2003).
6. Valente, T. W. Network interventions. Science 337, 49–53 (2012).
7. Granovetter, M. Threshold models of collective behavior. Am. J. Sociol. 78, 1116–1140 (1973).
8. Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small world’ networks. Nature 393, 440–442 (1998).
9. Burt, R. S. The network structure of social capital. Res. Organ. Behav. 22, 345–423 (2000).
10. Kitsak, M. et al. Identification of influential spreaders. Nat. Phys. 6, 888–893 (2010).
11. Moreno, F. & Makse, H. A. Influence maximization in complex networks through optimal percolation. Nature 524, 65–68 (2015).
12. Centola, D. & Macy, M. Complex contagions and the weakness of long ties. Am. J. Sociol. 113, 702–734 (2007).
13. Centola, D., Eguíluz, V. M. & Macy, M. W. Cascade dynamics of complex propagation. Phys. A 374, 449–456 (2007).
14. Romer, D. M., Meeder, B. & Kleinberg, J. (Eds.) Proc. 20th International Conference on World Wide Web (ACM, 2011).
15. Alatas, V., Banerjee, A., Chandrasekhar, A. G., Hanna, R. & Olken, B. A. Network structure and the aggregation of information: theory and evidence from Indonesia. Am. Econ. Rev. 106, 1663–1704 (2016).
16. Newman, M. E. & Park, J. Why social networks are different from other types of networks. Phys. Rev. E 68, 036123 (2003).
17. Newman, M. Networks: An Introduction (Oxford University Press, 2010).
18. WHO. Accelerating Work to Overcome the Global Impact of Neglected Tropical Diseases: A Roadmap for Implementation (WHO, Geneva, 2012).
19. Fenwick, A. et al. The schistosomiasis control initiative (SCI): rationale, development and implementation from 2002–2008. Parasitology 136, 1719–1730 (2009).
20. WHO. Investing to Overcome the Global Impact of Neglected Tropical Diseases: Third WHO Report on Neglected Tropical Diseases (WHO, 2015).
21. Fleming, F. M., Matovu, F., Hansen, K. S. & Webster, J. P. A mixed methods approach to evaluating community drug distributor performance in the control of neglected tropical diseases. Parasit. Vectors 9, 345 (2016).
22. Shuford, K. V., Turner, H. C. & Anderson, R. M. Compliance with anthelmintic drug administration under the programme to eliminate lymphatic filariasis. PLoS Negl. Trop. Dis. 7, e2447 (2013).
23. Babu, B. V. & Babu, G. R. Coverage of, and compliance with, mass drug administration under the programme to eliminate lymphatic filariasis in India: a systematic review. Trans. R. Soc. Trop. Med. Hyg. 108, 538–549 (2014).
24. Petters, L. E. & Woolridge, J. M. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. J. Appl. Econometr. 11, 619–632 (1996).
25. Cameron, A. C. & Trivedi, P. K. Microeconometrics Using Stata. 2nd edn, Vol. 5 (Stata Press, 2010).

Acknowledgements

We appreciate the long-term involvement of the people of our 17 study villages in Mayuge District, Uganda. Several field teams made this project possible: the laboratory technicians from the Vector Control Division in Uganda who collected the parasitology, the field assistants, in particular Benjamin Tinkitina, and local village surveyors who collected household data. The Schistosomiasis Control Initiative at Imperial College London and the Schistosomiasis Research Group at the University of Cambridge were instrumental to setting up this project and providing discussion. We thank Drs N.A.W. Bell, V. Thacker, and N. Loahakunakorn for helping prepare tablet computers. Financially supported by the Chancellor’s Fund of the University of Cambridge, the Schistosomiasis Control Initiative, the Wellcome Trust (Grants 083931/Z/07/Z and 100891/Z/13/Z), the Isaac Newton Trust (Grant 15.40(v)), and the Netherlands Organization for Scientific Research (Grant 452–04–333).
**Author contributions**
Conceived and designed the study: G.F.C. and A.A.K. Conducted the study and analyzed the data: G.F.C. Wrote the paper: G.F.C., A.A.K. and E.B. Contributed background materials/analysis tools/organization of field teams: G.F.C., A.F., E.B., A.A.K., N.B.K., E.M.T., and D.W.D.

**Additional information**
Supplementary Information accompanies this paper at doi:10.1038/s41467-017-01499-z.

**Competing interests:** A.F. is the executive director of the Schistosomiasis Control Initiative (SCI) and N.B.K. is employed by SCI. E.M.T. is the assistant commissioner of Health Services for the Vector Control Division of Ministry of Health and the National Coordinator for SCI of the Neglected Tropical Disease Control/Elimination Programme in Uganda. The remaining authors declare no competing financial interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017