INTERFACE

rsif.royalsocietypublishing.org

Research

CrossMark
click for updates

Author for correspondence:
Luis A. Nunes Amaral
e-mail: amaral@northwestern.edu

Social embeddedness in an online weight management programme is linked to greater weight loss

Julia Poncela-Casasnovas[1,2], Bonnie Spring[5], Daniel McClary[1,2], Arlen C. Moller[5,6], Rufaro Mukogo[1], Christine A. Pellegrini[5], Michael J. Coons[5], Miriam Davidson[5], Satyam Mukherjee[1,3] and Luis A. Nunes Amaral[1,2,3,4]

[1]Department of Chemical and Biological Engineering, [2]HHMI, [3]Northwestern Institute on Complex Systems, and [4]Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA
[5]Department of Preventive Medicine, Northwestern University, Chicago, IL 60611, USA
[6]Department of Psychology, Illinois Institute of Technology, Chicago, IL 60616, USA

The obesity epidemic is heightening chronic disease risk globally. Online weight management (OWM) communities could potentially promote weight loss among large numbers of people at low cost. Because little is known about the impact of these online communities, we examined the relationship between individual and social network variables, and weight loss in a large, international OWM programme. We studied the online activity and weight change of 22 419 members of an OWM system during a six-month period, focusing especially on the 2033 members with at least one friend within the community. Using Heckman's sample-selection procedure to account for potential selection bias and data censoring, we found that initial body mass index, adherence to self-monitoring and social networking were significantly correlated with weight loss. Remarkably, greater embeddedness in the network was the variable with the highest statistical significance in our model for weight loss. Average per cent weight loss at six months increased in a graded manner from 4.1% for non-networked members, to 5.2% for those with a few (two to nine) friends, to 6.8% for those connected to the giant component of the network, to 8.3% for those with high social embeddedness. Social networking within an OWM community, and particularly when highly embedded, may offer a potent, scalable way to curb the obesity epidemic and other disorders that could benefit from behavioural changes.

1. Introduction

Obesity has become a global health crisis, with over half a billion obese adults worldwide [1]. Obesity heightens the risk of chronic illnesses, including cardiovascular disease, diabetes and cancer [2,3]. Loss of 5% of initial body weight produces clinically meaningful improvement in risk biomarkers [4] and can be achieved through intensive lifestyle intervention [5,6]. This form of treatment entails delivery of multiple intervention sessions to encourage self-monitoring and modification of diet and physical activity and help patients' problem solve barriers to behavioural change [7,8]. It is well established that weight loss increases in proportion to the number of contacts with an interventionist and that reducing the total number of sessions to less than 12 over a six-month period compromises the efficacy of weight-loss treatment. Multiple treatment sessions are costly, however, and time-consuming for patients. More scalable, less burdensome treatment modalities are thus needed to reach the enormous population that needs help with weight loss.

Online weight management (OWM) programmes hold the potential to foster the spread of weight loss among large numbers of people at low cost [9–11]. Presently, at least 25 commercial online weight-loss programmes already offer social networking features (defined broadly as online tools that facilitate communication

THE ROYAL SOCIETY
PUBLISHING

between two members), but little is known about the impact of participation in these online communities. However, simulation studies conducted by Bahr et al. [12] suggest that dynamic changes within such networks might be leveraged for the purpose of helping to control weight.

Here, we studied the relationship between a set of variables and weight loss in one such OWM community. Our objective was to gain insight into individual and social network influences on weight management over time. Individual characteristics that predict greater weight loss in the in-person treatment context include higher initial weight, greater participation in treatment and greater self-monitoring of diet, activity and weight. We examined the impact of these variables alone and along with social network parameters. Whereas, the use of the social networking features of an OWM programme has been suggested to correlate with better weight-loss maintenance [13], little is known about how social networking in an OWM community might affect weight change. We examined two potential mechanisms that might be linked to weight change: (i) social contagion (note that 'social contagion' has a different meaning than 'physical contagion' in epidemiology, see [14–16]) and (ii) social support.

Christakis & Fowler's [17] analyses of data from the Framingham Heart Study [18] supported the social contagion hypothesis by showing that weight gain spreads throughout a social network and is better related to social than geographical distance. Critics of this work have argued that point estimates of this 'social network effect' are reduced and become statistically indistinguishable from zero after controlling for contextual effects [16]. Nevertheless, others have reported evidence supporting the social contagion of weight hypothesis, specifically among adolescents [19,20].

Further, Leahey et al. [21] found evidence for the spread of weight-loss motivation among the social contacts of overweight and obese young adults who were trying to lose weight. We hypothesize that weight loss may similarly spread by a process of contagion through online social space. Evidence of actual weight-loss contagion would emerge in this study if participants' weight loss could be predicted by the magnitude of weight loss among their online friends. The social support hypothesis posits that some other behaviour or activity of online friends, separate from their weight change, influences an individual to gain or lose weight. For example, provision of social support or persuasion by influential figures in the social network is another mechanism through which social networking might help participants lose weight.

Social support is well established to be an effective component of in-person weight-loss interventions [22] and it is important to understand if and how these social relationships play a role in the interventions initial success and maintenance. However, few studies of health behaviour interventions have explored the dynamic and emergent properties of social support networks within such interventions [23]. Importantly, it remains unknown whether all connections or friends are equally effective providers of weight-loss support [12]. If all friendships are equally potent sources of support, then the total number of friendships should predict the magnitude of weight loss. By contrast, if the degree of network embeddedness of the members is the key aspect of social networking, then this measure of embeddedness in the network—which can be quantified, for example, by the so-called k-shell index (see Material and methods for definition)—should have the greatest explanatory power.

The present research is significant because it is the first to investigate the relationship between individual and social network variables, and weight loss in an online community expressly focused on weight management, and to investigate which aspects of online social connectedness most strongly promote weight loss.

1.1. Data

We acquired de-identified data from a large international OWM programme.[1] The dataset includes sign-up date, age, height, gender and initial weight for 47 026 unique records. Additionally, the dataset included the timestamp and value of subsequent recorded weigh-ins and other types of recorded programme activity for the time period between 1 January 2009 and 31 December 2010. We define engagement time of a programme member as the number of days between his/her sign-up date and the date of his/her last recorded programme activity. We observe that of the 47 026 unique records in the dataset, only 22 419 are participating members of the OWM programme, which we define as providing evidence of returning to the system after their first visit. In table 1, we show summary statistics for different subpopulations in the dataset.

We distinguish among three types of recorded programme activity made available to us. The first type is recorded weigh-ins. We have access to the timestamp and value of each weight recorded. The second type is friendship requests. We have access to the direction and outcome of friendship requests, but not to the timestamps, which were not recorded by the system. The third type is online communication. It comprises the timestamps—but not the content—of posts or comments to forums and blogs, and messages to other programme members.

Although the dataset is unusually rich in many respects, limitations on other aspects impose some constraints on our analyses. First, a programme member may be engaged with the programme while not engaging in any activity that gets recorded—the individual may read blog and forum postings but not write any comments or post, or may weigh herself every day but never record a weigh-in. Such 'under-the-radar' activity will not be detectable. Second, we chose to restrict the term friendship to relationships that include the use of the 'friending' mechanism of the OWM programme (a friendship request followed by an acceptance) and do not consider other forms of communication as indicative of friendship. Third, because we do not know the time at which a friendship was established, we cannot attempt to estimate the effect of the duration of a friendship on outcomes. Finally, in order to control for the effect of time of engagement with the OWM programme, we restrict all analysis to each user's first six months in the system (see table 1 and figure 1 for details).

1.2. Selection bias and data censoring

In contrast to studies where participants are randomized and the social network is artificially created by the researchers [24], the OWM programme members form an organic, self-organizing community. We have merely observed their actions post-fact and have not intervened in the system in any way. For this reason, we must take special care in addressing the effects of self-selection and censoring [25].

To estimate weight change, we need at least two recorded weigh-ins. Because individuals with at least two weigh-ins

Table 1. Summary statistics for the different sets of members considered in this study. *Group 1*: all members with at least two weigh-ins ($N_{wi} \geq 2$); *Group 2*: members with at least two weigh-ins and engagement time of at least six months ($N_{wi} \geq 2$ and $t \geq 180$); *Group 3*: members with at least two weigh-ins and at least one friend ($N_{wi} \geq 2$ and $k \geq 1$); and *Group 4*: members with at least two weigh-ins, engagement time of at least six months and at least one friend ($N_{wi} \geq 2$ and $t \geq 180$ and $k \geq 1$). Note that *Group 4* $\subset$ *Group2* $\subset$ *Group 1* and also *Group 3* $\subset$ *Group 2*. Values for online communications and per cent weight change were measured at 180 days of engagement.

|		Group 1 ($N_{wi} \geq 2$)	Group 2 ($N_{wi} \geq 2$ and $t \geq 180$)	Group 3 ($N_{wi} \geq 2$ and $k \geq 1$)	Group 4 ($N_{wi} \geq 2$ and $t \geq 180$ and $k \geq 1$)
set size		22 419	5409	2033	842
age	mean	43	44	41	42
	s.d.	15	17	12	12
	10th percentile	28	29	26	27
	90th percentile	59	60	57	58
gender	% female	83	82	88	88
initial BMI (kg m^{-2})	mean	31.5	31.6	33.2	33.0
	s.d.	7.2	7.2	7.8	7.4
	10th percentile	23.6	23.8	24.5	24.7
	90th percentile	41.1	41.2	43.8	43.6
engagement time (days)	mean	147	307	214	331
	s.d.	152	152	152	152
	10th percentile	6	201	23	183
	90th percentile	358	495	428	552
number of weigh-ins	mean	10	21	16	25
	s.d.	14	21	18	23
	10th percentile	2	6	3	7
	90th percentile	22	36	31	43
online communications (events/week)	mean	0.23	0.7	2.36	4.3
	s.d.	2.47	4.7	7.86	11.2
	10th percentile	0	0	0	0.04
	90th percentile	0.12	0.46	4.83	11.7
percentage weight change	mean	−2.95	−4.76	−4.64	−6.78
	s.d.	4.96	6.25	5.9	6.83
	10th percentile	−8.45	−13.1	−12.55	−16.2
	90th percentile	1.15	1.65	0.98	0.57

represent a self-selected, potentially biased sample, we analysed the data using the sample-selection procedure developed by Heckman and others [25] (see Material and methods for details).

Another possible source of selection bias could be the establishment of friendships. We thus also apply Heckman's method to correct for this constraint. Finally, not all programme members with at least two recorded weigh-ins have engagement times long enough for us to estimate the effect of their programme participation on weight change (figure 1). Consequently, we set a threshold of six months of engagement time with the OWM programme as requirement for estimating the effect of OWM programme participation on outcomes (table 1), and we correct for the probability of staying in the system for at least that long. Although six months is traditionally considered the standard duration time needed to determine whether a weight loss intervention has been

successful [26,27], we also test the robustness of our results by repeating our analysis setting a threshold of a minimum of four months of engagement with the programme.

2. Results and discussion

We found that the 47 026 unique records can be classified into three distinct populations (figure 1): 40% of the individuals never return to the site after the first visit, 3% of members stay in the programme for a short amount of time (i.e. approx. 17 days, signifying high attrition rate) and 47% of members stay in the programme for a longer period of time (approx. 214 days, signifying low attrition rate). When focusing only on the programme members with at least one friend, we observed that the population (2033 programme members) is made almost entirely (96%) of low attrition
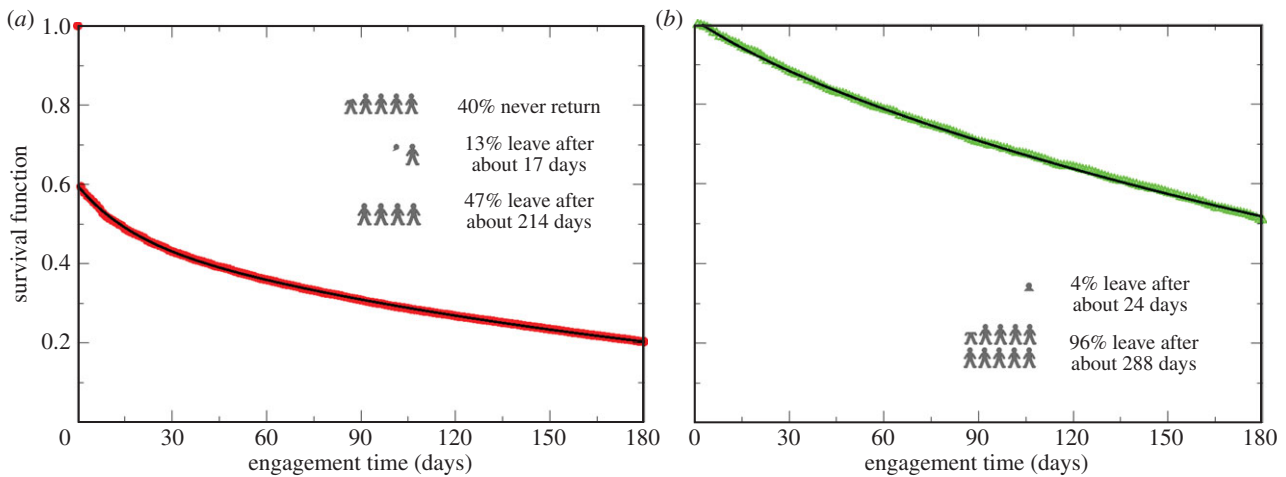
Figure 1. Cumulative distribution of engagement time for all unique identifiers in the database (*a*) and for programme members with at least one friend (*b*), and their corresponding fits to a linear superposition of exponentials ($Y = 0.47\exp(-t/214) + 0.13\exp(-t/17)$ and $Y = 0.96\exp(-t/288) + 0.04\exp(-t/24)$, respectively). We defined engagement time as the number of days between signing up to the programme and the last day of recorded activity. We restrict the plotting to 180 days because our study focuses on that period of time. Note the extraordinarily good agreement between data (red or green symbols) and fit (black line). (Online version in colour.)

programme members. Thus, engagement with social network tools is correlated with longer engagement times.

2.1. Weight loss outcome

We computed the average percentage body weight change δ for the 5409 programme members with at least two recorded weigh-ins and for whom we have visible OWM programme engagement data until the six-month mark. We also built the friendship network of the OWM programme members (figure 2). Surprisingly, we found that 91% of programme members with at least two weigh-ins, whom we denote 'isolated', did not establish online friendships within the OWM system during the observation period. Among the 2033 programme members that did establish online friendships, nearly 74% aggregate into a single cluster, the so-called 'giant component' (or GC, is a subgraph of nodes connected to each other that constitutes the majority or largest fraction of the network, see Material and methods), whereas the remaining 26% formed small clusters comprised typically of two or three programme members, with a few of size four and five, a single cluster of six and another one of nine individuals.

2.2. Online activity

Not surprisingly, we found that member recorded online communications increase dramatically with level of embeddedness in the friendship network. Isolated members engaged in the least recorded online communications (average of 0.04 events per week), whereas members belonging to small components had intermediate levels of recorded online communications (average of 0.4 events per week) and members in the GC had the largest recorded online communications (average of 5.05 events per week).

2.3. Modelling weight change

We investigate a linear regression model for percentage weight change that takes into account known factors affecting outcomes,

$$\delta(i) = A_0 + A_1 I(i) + A_2 N_{wi}(i) + A_3 t(i) + A_4 C(i)$$
$$+ A_5 S(i) + \epsilon(i), \qquad (2.1)$$

where $A_0 - A_5$ are the coefficients we want to estimate, $I(i)$ is the initial body mass index (BMI) [21], $N_{wi}(i)$ is the number of recorded weigh-ins, which we use as a proxy for self-monitoring adherence [28,29], $t(i)$ is the engagement time (measured from enrolment day to the last day of recorded activity for a give user), $C(i)$ is the total number of recorded online communications, $S(i)$ is a quantification of social engagement and $\epsilon(i)$ accounts for the many unobserved determinants of weight change.

Several aspects of social networking could plausibly impact weight loss [13,30–32], and a multitude of measures could be used to quantify those aspects. As we do not *a priori* know which factors do influence weight loss, we systematically investigated the two most plausible hypotheses about social networks' effectiveness. The first hypothesis is that friendships within an OWM programme social network lead to peer-to-peer 'social contagion' of weight change outcomes [17]. Interestingly, we found that an OWM member's weight change is significantly correlated with the average weight change of his/her friends *only if that is the single social networking factor included in the model*. However, the effect of friends' weight change is not statistically significant or is barely above the significant level when other social networking factors are considered (see tables 2 and 3 for details). We then can rule out any important correlation between the average weight change of an individual's group of friends and his or her own change for the individuals in our dataset and focus from now on in other possible explanations. The second hypothesis is that social networking provides a supportive environment for programme members to become more successful at achieving their weight management goals. Under the second hypothesis, social networking helps programme members maintain motivation for behaviours that enable them to lose weight and to maintain weight loss.

We next discuss different proxies for expressing the manner in which support could be mediated through the online social network: (i) degree or total number of friendships, (ii) *k*-shell membership [33], and (iii) betweenness centrality. The *k*-shell of a network is the set of all nodes belonging to the *k* core of that network (that is, the maximal subgraph of that network having minimum degree of at least *k*), but not to the $(k + 1)$ core. The *k*-shell index measures the embeddedness of a node
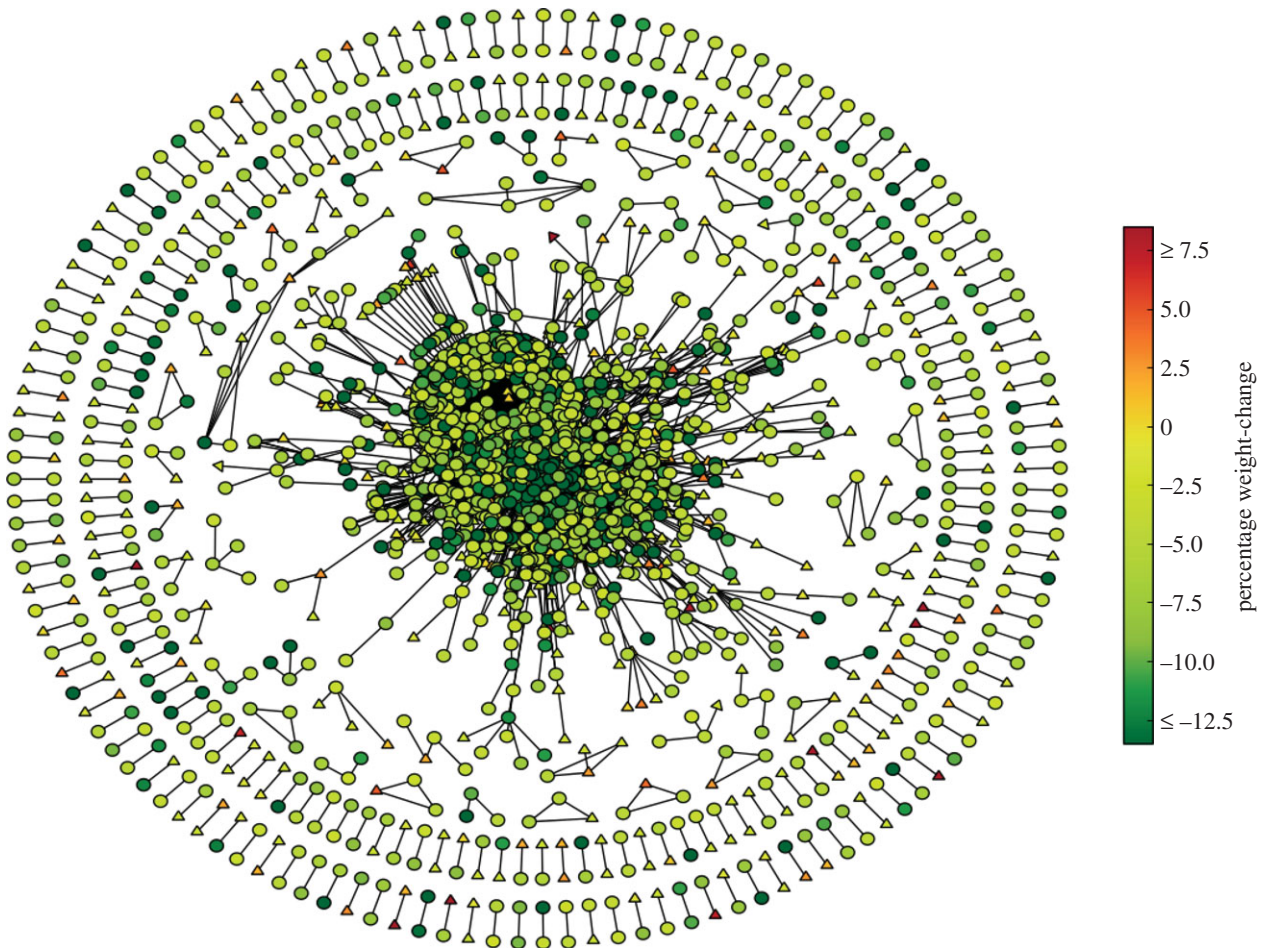
Figure 2. Friendship connections among programme members. We denote members who lose weight by circles, and those who gain weight by triangles. We colour the symbols according to total per cent weight change. The vast majority of networked members (74%) belong to a single cluster, the so-called 'giant component'. Fifty-four per cent of the members in the giant component lost 5% or more of their baseline body weight by their sixth month in the programme.

within the network (the higher the index, the more embedded the node is in the structure). On the other hand, the betweenness of a node is one of the standard measures for node centrality. It measures the fraction of shortest paths in the network that go through that given node (the higher this fraction, the more central is the node). Note that these three network measures can be correlated, but describe in general different aspects of the network structure (see Material and methods for more details).

To find the best possible model for weight change given the available data, we proceeded as follows. First, we examined the contribution of each variable independently to determine whether or not the variable should be included. (If a variable was not significant on its own, it would not be included in the final model.) Next, to assess how many of the (individually) significant variables to add to the model, we considered the inherent trade-off between model complexity and statistical explanatory power. That is to say, we aimed to find the simplest model with the maximum statistical explanatory power. On the other hand, for more details on the units and interpretation of the coefficients in the model, see Material and methods.

We found that betweenness centrality was not statistically significant in our weight loss model (not even when considered as the only network variable). When considered independently, the number of friends appeared to significantly correlate with weight change outcomes. However, when testing all proxies

together in a single model, we found that the k-shell index was the factor with the greatest explanatory power and greatest significance and that the explanatory power of the model could not be increased by adding other variables (see table 2 for a detailed comparison of all these models).

Model parameter estimation confirms that initial BMI and total number of weigh-ins have significant explanatory power, being correlated with greater weight loss. The engagement time of the users is also significant (note that, although our model aims to measure weight change at 180 days, not all users necessarily have a last weigh-in on that particular day, so we actually consider the closest one around that day, hence, considering their particular engagement time as a variable). Although we included subject age in all of our initial models, age was never significantly correlated with weight change. A likely reason is that our population is comprised largely women in middle age (table 1). Hence, the range of ages available in the sample may have been too restricted to detect the effect of age. On the other hand, the number of online communications is also a significant predictor for weight change at 180 days and at 120 days, although its level of statistical significance always decreases when considered along with other network variables, as opposed to on its own. This can be due to the fact that online communication acts as a proxy for peer interaction (in general, the more people a user knows in the system, or the more central she is in the network, the more messages she will send and receive, see table 2). The

Table 2. Coefficients [s.e.] for the linear models tested for percentage weight change. Every row corresponds to a model for weight change of individuals with an engagement time of 180 days and at least one friend, after correcting for self-selection bias. The dataset used for this analysis is *Group 1*, where the total number of observations is 22 419 (842 uncensored), except for models that include as variable the average weight change of a user's friends (822 uncensored observations). We denote as uncensored observations those for whom we have the values of all variables, and as censored observations those for whom we are missing the values of some variables. Superscript n.s. denotes not significant.

intercept	initial BMI	number of weigh-ins	time in the system	online communic.	k-shell index	number of friends	betweenness	avg. weight change friends	inv. Mills ratio	adj. R^2 (%)
0.35$^{n.s.}$ [1.88]	−0.14*** [0.03]	−0.09*** [0.01]	−0.009*** [0.001]						1.40* [0.58]	20.8
3.84* [1.91]	−0.13*** [0.03]	−0.08*** [0.01]	−0.009*** [0.001]	−0.005*** [0.001]					−0.35$^{n.s.}$ [0.63]	24.1
4.48$^{n.s.}$ [1.87]	−0.12*** [0.03]	−0.09*** [0.01]	−0.008*** [0.001]		−0.56*** [0.07]				−0.44$^{n.s.}$ [0.59]	26.5
0.32$^{n.s.}$ [1.87]	−0.13*** [0.03]	−0.09*** [0.01]	−0.009*** [0.002]			−0.014** [0.005]			1.31* [0.57]	21.4
0.26$^{n.s.}$ [1.88]	−0.13*** [0.03]	−0.09*** [0.01]	−0.009*** [0.002]				−17.37$^{n.s.}$ [13.55]		1.40* [0.58]	20.9
0.78$^{n.s.}$ [1.90]	−0.12*** [0.03]	−0.09*** [0.01]	−0.010*** [0.002]					0.04* [0.01]	1.42* [0.57]	21.2
5.41$^{n.s.}$ [1.92]	−0.12*** [0.03]	−0.08*** [0.01]	−0.008*** [0.001]	−0.002** [0.001]	−0.43*** [0.08]	−0.002$^{n.s.}$ [0.005]		0.02$^{n.s.}$ [0.01]	−0.83$^{n.s.}$ [0.63]	27.0
5.35** [1.90]	−0.13*** [0.03]	−0.09*** [0.01]	−0.008*** [0.001]	−0.002* [0.001]	−0.46*** [0.08]				−0.92$^{n.s.}$ [0.63]	27.0

***$p < 0.001$.

**$p < 0.01$.

*$p < 0.05$.

Table 3. Coefficients [s.e.] for the linear models tested for percentage weight change. Every row corresponds to a model for weight change of individuals with an engagement time of 120 days and at least one friend, after correcting for self-selection bias. The total number of observations is 22 419 (1104 uncensored), except for models that include as variable the average weight change of a user's friends (1076 uncensored observations). We denote as uncensored observations those for whom we have the values of all variables, and as censored observations those for whom we are missing the values of some variables. Superscript n.s. denotes not significant.

intercept	initial BMI	number of weigh-ins	time in the system	online communic.	k-shell index	number of friends	betweenness	avg. weight change friends	inv. Mills ratio	adj. R^2 (%)
−2.02$^{n.s.}$ [1.31]	−0.11*** [0.02]	−0.09*** [0.01]	−0.007*** [0.001]						1.87*** [0.40]	17.8
0.10$^{n.s.}$ [1.41]	−0.11*** [0.02]	−0.09*** [0.01]	−0.006*** [0.001]	−0.004*** [0.001]					0.86$^{n.s.}$ [0.48]	18.7
0.04$^{n.s.}$ [1.34]	−0.10*** [0.02]	−0.09*** [0.01]	−0.006*** [0.001]		−0.31*** [0.06]				0.92$^{n.s.}$ [0.43]	19.7
−2.01$^{n.s.}$ [1.31]	−0.10*** [0.02]	−0.09*** [0.01]	−0.007*** [0.001]			−0.008$^{n.s.}$ [0.005]			1.82*** [0.40]	17.9
−2.04$^{n.s.}$ [1.31]	−0.10*** [0.02]	−0.09*** [0.01]	−0.007*** [0.001]				−6.03$^{n.s.}$ [11.91]		1.87*** [0.40]	17.7
−1.54$^{n.s.}$ [1.27]	−0.08*** [0.02]	−0.10*** [0.01]	−0.007*** [0.001]					0.03** [0.01]	1.61* [0.38]	18.8
1.55$^{n.s.}$ [1.36]	−0.08*** [0.02]	−0.09*** [0.01]	−0.006*** [0.001]	−0.003* [0.001]	−0.27*** [0.06]			0.03* [0.01]	0.10$^{n.s.}$ [0.46]	21.4

***$p < 0.001$.

**$p < 0.01$.

*$p < 0.05$.

degree of embeddedness in the social network (k-shell) is always strongly correlated with increasing weight loss. The average weight change of a user's friends is not a significant variable for the model at 180 days whenever other network variables are also included, and only minimally significant for the model at 120 days.

After correcting for selection bias (see Material and methods for details), and defining the network variable as the k-shell index, $S(i) = K(i)$, the variables in equation (2.1) account for approximately 27% of the observed variance of the data (table 2). This value is quite large for a social system, given the fact that there are a large number of other factors we do not have information about, such as diet, exercise level, education level, motivation and so on. For example, Gallos et al. [34] identified strong long-range spatial clustering of obesity, and that such clustering mirrors the clustering observed in the strength of economic activity related to food production and sale.

Moreover, to test the robustness of these results, we repeat the same analysis for weight change measured at 120 days (table 3), observing that the results are very consistent with the ones found for 180 days. For comparison reasons, we also show in table 4 a selection of the best models (highest adjusted R^2 and least possible number of variables) for percentage weight change at 120 and 180 days, with and without social network variables.

In our estimation, we correct for the probability of staying in the programme for at least 180 (or 120) days and for the probability of having at least one friend. Whenever the inverse Mills Ratio (or the ratio of the probability density function to the cumulative distribution function of a distribution, see Material and methods for details) is significant, it means that self-selection introduces a bias in the sample that needed to be corrected by the Heckman procedure.

It is worth mentioning that when considering a linear model *without* the social networking effect, but keeping the other individual variables, the explanatory power always drops (from around 27% to around 20% for the model at 180 days, and from around 21% to around 18% for the model at 120 days, see table 4), demonstrating the importance of social embeddedness on weight loss in online communities.

Similarly, we note that the explanatory power is higher for the set of models that correct for both the probability of having at least one friend and the probability of remaining in the system for a given amount of time (120 days or 180 days), as compared to those that only correct for time in the system but not for having friends (compare models in tables 4 and 5, respectively). Despite this, however, the results for the two sets of models are otherwise consistent with each other in terms of the contribution and statistical significance of the different variables.

It is also worth mentioning that we systematically investigated the possible impact of including nonlinear second-order terms in the models for weight loss, but we consistently found that the increase in explanatory power given by the new adjusted R^2 was typically less than a 1% (or nothing at all). In most cases, the crossed terms were not significant. Given the trade-off between a large increase in complexity of the models and a small increase in their explanatory power compare with a linear one, we decided to present results just for linear models.
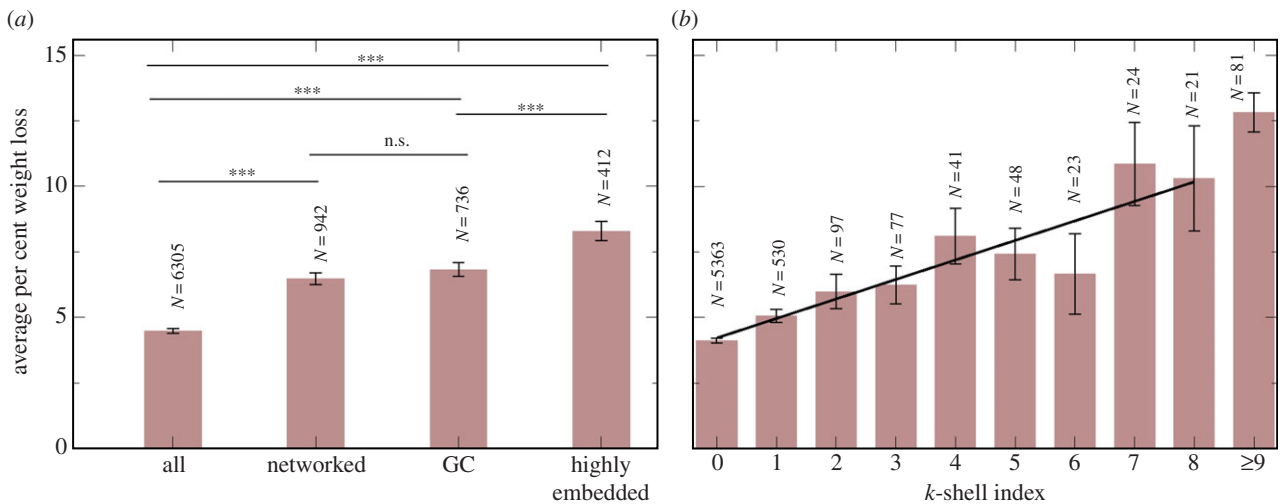
Table 4. Coefficients [s.e.] for the best linear models proposed for percentage weight change of the set of programme members with and without social networking variables. Every row corresponds to a model for weight change for individuals with an engagement time of 120 or 180 days and at least one friend, after correcting for self-selection bias. The total number of members is 22 419 (1104 uncensored observations at 120 days and 842 uncensored observations at 180 days). We denote as uncensored observations those for whom we have the values of all variables, and as censored observations those for whom we are missing the values of some variables. u.o. denotes uncensored observations.

	intercept	initial BMI	number of weigh-ins	time in the system	online communications	k-shell index friend	avg. weight change	inv. Mills ratio	adj. R^2 (%)
weight loss at 180	0.35[n.s.] [1.88]	−0.14*** [0.03]	−0.09*** [0.01]	−0.009*** [0.001]				1.40* [0.58]	20.8
days (842 u.o.)	5.35** [1.90]	−0.13*** [0.03]	−0.09*** [0.01]	−0.008*** [0.001]	−0.002* [0.001]	−0.46*** [0.08]		−0.92[n.s.] [0.63]	27.0
weight loss at 120	−2.02[n.s.] [1.31]	−0.11*** [0.02]	−0.09*** [0.01]	−0.007*** [0.001]				1.87*** [0.40]	17.8
days (1104 u.o.)	1.55[n.s.] [1.36]	−0.08*** [0.02]	−0.09*** [0.01]	−0.006*** [0.001]	−0.003* [0.001]	−0.27***[0.06]	0.03*[0.01]	0.10[n.s.] [0.46]	21.4

***$p < 0.001$.
**$p < 0.01$.
*$p < 0.05$.

Figure 3. Engagement in online social networking correlates with incremental weight loss at six months. (a) The networked members belonging to the giant component (and specially, the ones highly embedded in the network—k-shell higher or equal than 2) achieved the greatest weight loss. (b) Percentage weight loss increases by 0.7% for every unit increase in k-shell index (corr. coeff. = 0.91). Note that non-networked members have $K(i) = 0$. Bars represent standard errors, asterisks (***) denote $p < 0.001$ and n.s. denotes not significant. (Online version in colour.)

Table 5. Coefficients [s.e.] for the best linear models proposed for percentage weight change of the set of programme members. Every row corresponds to a model for weight change for individuals with an engagement time of 120 or 180 days, after correcting for self-selection bias. The total number of members is 22 419 (7383 uncensored observations at 120 days and 5409 uncensored observations at 180 days). We denote as uncensored observations those for whom we have the values of all variables, and as censored observations those for whom we are missing the values of some variables. Superscript n.s. denotes not significant. u.o. denotes uncensored observations.

	intercept	initial BMI	number of weigh-ins	time is the system	online communications	inv. Mills ratio	adj. R^2 (%)
weight loss at	$0.45^{n.s.}$	-0.18***	-0.06***	-0.006***		2.99***	12.8
180 days	[1.16]	[0.01]	[0.01]	[0.001]		[0.75]	
(5409 u.o.)	$2.07^{n.s.}$	-0.18***	-0.06***	-0.006***	-0.006*** [0.001]	1.61*	13.9
	[1.15]	[0.01]	[0.01]	[0.001]		[0.75]	
weight loss at	-5.96***	-0.12***	-0.03***	-0.0043***		6.73***	12.8
120 days	[0.87]	[0.01]	[0.01]	[0.0004]		[0.63]	
(7383 u.o.)	-4.90***	-0.12***	-0.04***	-0.0042***	-0.005*** [0.001]	5.78***	13.1
	[0.84]	[0.01]	[0.01]	[0.0004]		[0.62]	

***$p < 0.001$.
**$p < 0.01$.
*$p < 0.05$.

Finally, we compared our weight loss results with those obtained for a null model where the subjects are reshuffled within the network structure, but their individual attributes (such as initial BMI, weight change or number of weigh-ins) are preserved. As one would expect, we found that in this null model none of the network variables significantly correlated with weight loss anymore, whereas the individual variables remained significant, contributing similarly in terms of the value and significance of the coefficients, and rendering an overall explanatory power very consistent with the original model *without* network variables (see Material and methods for more details).

Next, and to show graphically the main results found in our analysis, we plot in figure 3 the explicit correlation between weight loss and social networking engagement

and, specifically, the k-shell index. We found that the members belonging to the GC lose around 6.8% of their body weight in six months, a value above the clinically significant threshold (5%), and also higher than the value for the general population (4.5%) or the non-networked members (4.1%). Note that the differences in weight loss for all pairs of sets shown are statistically significant (except for networked programme members versus programme members belonging to the GC). Moreover, the set of members who are highly embedded in the network, $K(i) \geq 2$, lose even more weight (around 8.3%, figure 3a).

Our most remarkable finding, however, is the incremental effect of the degree of network embeddedness, as measured by having an increasing k-shell index (figure 3b). While this correlation cannot establish a causal relationship, it nonetheless

Table 6. Results for the linear model and probit selection for weight change of individuals with an engagement time of 180 days and at least one friend, after correcting for self-selection bias. The social networking variable used is the k-shell index. We denote as uncensored observations those for whom we have the values of all variables, and as censored observations those for whom we are missing the values of some variables. We mark in italic the coefficients that are statistically significant (p-value < 0.001).

model	coef.	s.e.	*p*-value
weight change (adjusted for self-selection)			
intercept	4.48	1.87	0.02
initial BMI (I)	*−0.12*	0.03	<0.00001
number of weigh-ins (N_{wi})	*−0.09*	0.01	<0.00001
time in the system (t)	*−0.008*	0.001	<0.00001
social engagement (S)	*−0.56*	0.07	<0.00001
probit selection			
intercept	*−2.30*	0.07	<0.00001
initial BMI (I)	*0.009*	0.002	<0.00001
weight change by day 20 (δ_{20})	*−0.03*	0.01	<0.00001
online communications by day 20 (C_{20})	*0.021*	0.001	<0.00001
number of weigh-ins at day 20 ($N_{wi,20}$)	*0.024*	0.004	<0.00001
inverse Mills ratio	−0.44	0.59	0.46
total number of observations		22 419	
number of censored observations		21 577	
number of uncensored observations		842	
adjusted R^2		26.5%	

strongly suggests the positive relation between embeddedness in a social network and weight loss, and provides insight into the mechanisms by which social networks may influence outcomes. Moreover, these results are consistent with previous findings in other related fields such as the study of the spreading of information or trends online, highlighting the importance of high k-shell individuals [33,35–37] in the process. While we acknowledge the limitations of our data, namely that is not a representative sample of the US population, it does represent the online population interested in weight management, which is a population that is likely to grow in the future.

Indeed, our study provides the first rigorous demonstration that embeddedness in an online social network of individuals aiming to control their weight correlates with members' weight loss, above and beyond the benefits of adherence to self-monitoring or simply having friends *per se*. Given the scalability of online interventions and potential benefits of virtual social networks, it is essential to continue exploring and optimizing this important new tool in the arsenal of public health interventions.

3. Material and methods

3.1. Body mass index definition
BMI is a measure of human body fat based on an individual's weight and height: $\text{BMI} = \text{weight(lbs)}\cdot 703/(\text{height(in)})^2$. Using this index, one can define an overweight individual as someone with $25 < \text{BMI} \le 30$, and an obese person as someone with $\text{BMI} > 30$.

3.2. Correcting for self-selection bias
To better ascertain the effect of online friendships on weight loss, we need to restrict our analysis only to those individuals with an engagement time of six months and at least one friend. Nonetheless, by doing so, we could be introducing some bias in our analysis. In order to account for this possible selection bias and censored data, we follow the work developed by Heckman and others [25]. This is a two-step method where we first use the entire dataset to estimate the probability of having a data point with those particular restrictions (having at least one friend and an engagement time of at least six months), and then we apply that conditional probability as a correction in the linear model presented in equation (2.1). The variables used in the first step should be different from the ones in the second step, the linear model. In our case, for the first step, we parametrize a linear probit selection model for the probability of having an engagement time of at least six months (180 days) and having at least one friend given by

$$\text{Prob}(t_i \ge 180 \wedge k_i > 0) = D_0 + D_1 I(i) + D_2 N_{wi,20}(i) + D_3 C_{20}(i) + D_4 \delta_{20}(i) + v(i), \quad (3.1)$$

where $D_0 - D_4$ are the coefficients we want to estimate, $I(i)$ is the initial BMI, $N_{wi,20}(i)$ is the number of recorded weigh-ins during the first 20 days of engagement, $C_{20}(i)$ is the total number of recorded online communications during the first 20 days of engagement, $\delta_{20}(i)$ is the total per cent weight change recorded by day 20 and $v(i)$ accounts for the many unobserved determinants of an engagement time of at least six months. As we mentioned before, we repeated this analysis for all models, for the case of an engagement time of four and six months. In table 6, we show as an example, the complete results of the linear model and the probit selection for the weight change at six months.

From the probit, we found that the probability of having at least one friend and an engagement time of at least four

10

Table 7. Correlations between social network variables used in the weight change models for *Group 4*. Between every pair of variables, we show the value of the coefficient for the corresponding linear regression between them (the values in the first column are always the dependent variables), the standard error and the adjusted R^2 of that particular model. Superscript n.s. denotes not significant.

	online communications	k-shell index	log(betweenness)
k-shell index	0.0080***		
	[0.0002]		
	33.4%		
log(betweenness)	0.0020***	0.23***	
	[0.0003]	[0.03]	
	7.0%	6.5%	
numbers of friends	0.028***	3.24***	3.39***
	[0.003]	[0.19]	[0.58]
	4.8%	10.8%	45.1%

***$p < 0.001$.
**$p < 0.01$.
*$p < 0.05$.

months increases with increasing initial BMI, the number of communications and number of weigh-ins during the first 20 days in the system, and decreases with increasing per cent weight change by day 20. Whenever the coefficient of the inverse Mills ratio is not significant, we can conclude that there was no selection bias affecting significantly our estimates of the parameters of equation (2.1) in that particular model.

Inverse Mills ratio is the ratio of the probability density function to the cumulative distribution function of a distribution. It can be used in regression analysis to take account of a possible selection bias, and it is based on the fact that, if a dependent variable is censored (that is, not for all subjects there is an observed value for it), it causes a concentration of observations at zero values, and this violates the assumption of zero correlation between independent variables and the error term.

3.3. Units of the variables in the model

In all analyses performed in this study, BMI is measured in kilograms per square metre, number of weigh-ins is dimensionless, time in the system is measured in days, number of online communications is dimensionless, *k*-shell is dimensionless, degree is dimensionless, betweenness is dimensionless, the average weight change of a users' set of friends is quantified as a fraction so is dimensionless.

Naturally, the value of the coefficients estimated for the different models depends on the units the variables are measured in (for example, we could measure the degree of the individuals in tens of friends, instead of number of friends). A change in units would result in a simple scaling of the value of the coefficient, but the amount of variability in the data explained by that variable would not change at all.

3.4. Network theory: glossary and analysis

In order to analyse the effects of social networking on weight loss outcomes in the OWM programme, we construct a network of the friendships between programme members (figure 2). We build the network from a list of pairs of de-identified identifiers that represent pairs of programme members who have an online friendship (defined as a friendship request followed by acceptance). Next, we define some network terms that are used in this study. More information on networks theory can be found in reviews such as [38,39].

Degree is the number of edges connected to a node. In the case of a friendship network, it would be the number of friends an individual has.

GC is a subgraph of nodes that are connected to each other that constitutes the majority or largest fraction of the network.

Isolated components are small groups of nodes that are disconnected from the GC in the network.

Betweenness of a node is one of the standard measures for node centrality. It measures the fraction of shortest paths in the network that go through that given node (the higher this fraction, the more central is the node).

k-core of a network is the maximal subgraph of that network having minimum degree at least k.

k-shell of a network is the set of all nodes belonging to the k-core of that network, but not to the $(k + 1)$ core. The k-shell index measures the embeddedness of a node within the network (the higher the index, the more embedded the node is in the structure).

Calculation of k-shell decomposition. The k-shell index of the nodes in a network is calculated using the following algorithm [33,40]: take the entire network and evaluate the degree of each node. Then remove all nodes with degree 1. Once we are done, we repeat the process, evaluating and pruning all nodes again until there are no more nodes left with degree 1. Then we can identify and label the nodes with k-shell index $K = 1$ as all those nodes pruned during the iterative process. Next, we evaluate and recursively prune all nodes with degree 2, to get the nodes with k-shell index $K = 2$, and so on. The k-shell index measures hierarchy and level of embeddedness of a node within the network (the higher the index, the more embedded the node is in the structure). Note that, because of the iterative procedure that the algorithm is based on, k-shell and degree do not coincide, in general, for a given node.

Correlations between network metrics. It is important to note that, even when different network metrics might be correlated sometimes (table 7), they describe different aspects of the network structure. Thus, for example, a node with high degree can also have high betweenness and high k-shell, but a counter example would be a situation where a node is the centre of a star-like neighbourhood that is not in the core of the network, but in the periphery: this node would have high degree but low betweenness and low k-shell index. Similarly, betweenness and k-shell might be correlated, but not necessarily. A counter example would be a node that by a few links connects two otherwise independent communities (that is, it acts like a bridge): this

Table 8. Coefficients [s.e.] for the linear null models for percentage weight change. Every row corresponds to a model for weight change of individuals with an engagement time of 180 days and at least one friend, after correcting for self-selection bias. In these null models, the individuals have been reshuffled to different locations in the network, but preserving their personal features. The dataset used for this analysis is *Group 1*, where the total number of observations is 22 419 (842 uncensored), except for models that include as variable the average weight change of a user's friends (822 uncensored observations). We denote as uncensored observations those for whom we have the values of all variables and as censored observations those for whom we are missing the values of some variables. Superscript n.s. denotes not significant.

intercept	initial BMI	number of weigh-ins	time in the system	online communic.	k-shell index	number of friends	betweenness	inv. Mills ratio	adj. R^2 (%)
3.85*	−0.13***	−0.08***	−0.009***	−0.005***				−0.35 n.s.	24.1
[1.91]	[0.03]	[0.01]	[0.001]	[0.001]				[0.63]	
0.34 n.s.	−0.13***	−0.09***	−0.009***		0.004 n.s.			1.39*	20.7
[1.89]	[0.03]	[0.01]	[0.002]		[0.080]			[0.58]	
0.39 n.s.	−0.14***	−0.09***	−0.009***			−0.004 n.s.		1.40*	20.8
[1.88]	[0.03]	[0.01]	[0.002]			[0.005]		[0.57]	
0.38 n.s.	−0.14***	−0.09***	−0.010***				11.31 n.s.	1.40***	20.8
[1.88]	[0.03]	[0.01]	[0.001]				[13.57]	[0.58]	

***$p < 0.001$.

**$p < 0.01$.

*$p < 0.05$.

Table 9. Lack of correlations between number of online communications and other social network variables used in the weight change null models for *Group 4*. Between every pair of variables, we show the value of the coefficient for the corresponding linear regression between them (the values in the first column are always the dependent variables), the s.e. and the adjusted R^2 of that particular model. Superscript n.s. denotes not significant.

	online communications
k-shell index	5.3×10^{-6} n.s.
	$[2.7 \times 10^{-4}]$
	−0.05%
log(betweenness)	−0.0003 n.s.
	[0.0005]
	−0.1%
numbers of friends	−0.001 n.s.
	[0.002]
	−0.04%

***$p < 0.001$.

**$p < 0.01$.

*$p < 0.05$.

node would have very high betweenness, but very low k-shell index and degree. Finally, we acknowledge the possibility of finding correlations between degree of a user and her amount of online communication.

3.5. Null model

In order to strengthen the results presented in this paper, we compare them with those obtained with a null model, the subjects are reshuffled within the network structure, but preserving all their individual attributes (such as initial BMI, weight change or number of weigh-ins). We found that in this null model, none of the network variables are significantly correlated with weight loss anymore. We present in table 8 the results for one of the 100 independent realizations of the reshuffling process, the results being highly consistent between all realizations. We observe that the individual variables, such as initial BMI, number of weigh-ins or number of online communications are still significantly correlated with weight loss, and they contribute similarly in terms of the value and significance of the coefficients. The overall explanatory power of the null model is very consistent with that of the corresponding original model *without* network variables. This is true both for the model when measuring weight loss at 180 days and 120 days. We also observe that in the null model, the number of online communications and all network variables are no longer correlated (table 9).

Endnote

[1]The dataset is available upon request.

References

1. World Health Organization, and 2011 Obesity. WHO See http://www.who.int/topics/obesity/en/ (accessed 5 May 2012).

2. Bays HE, Chapman RH, Grandy S. 2007 The relationship of body mass index to diabetes mellitus, hypertension and dyslipidaemia: comparison of data from two national surveys. *Int. J. Clin. Pract.* **61**, 737–747. (doi:10.1111/j.1742-1241.2007.01336.x)

3. Field AE, Coakley EH, Must A, Spadano JL, Laird N, Dietz WH, Rimm E, Colditz GA. 2001 Impact of overweight on the risk of developing common chronic diseases during a 10-year period. *Archiv. Intern. Med.* **161**, 1581–1586. (doi:10.1001/archinte.161.13.1581)

4. Franz MJ, VanWormer JJ, Crain AL, Boucher JL, Histon T, Caplan W, Bowman JD, Pronk NP. 2007 Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. *J. Am. Diet. Assoc.* **107**, 1755–1767. (doi:10.1016/j.jada.2007.07.017)

5. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM. 2002 Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* **346**, 393–403. (doi:10.1056/NEJMoa012512)

6. Pi-Sunyer X. 2007 Reduction in weight and cardiovascular disease risk factors in individuals with type 2 diabetes. *Diab. Care* **30**, 1374–1383. (doi:10.2337/dc07-0048)

7. Appel LJ *et al.* 2011 Comparative effectiveness of weight-loss interventions in clinical practice. *N. Engl. J. Med.* **365**, 1959–1968. (doi:10.1056/NEJMoa1108660)

8. The Diabetes Prevention Program. 1999 Design and methods for a clinical trial in the prevention of type 2 diabetes. *Diab. Care* **22**, 623–634. (doi:10.2337/diacare.22.4.623)

9. Gold BC, Burke S, Pintauro S, Buzzell P, Harvey-Berino J. 2007 Weight loss on the web: a pilot study comparing a structured behavioral intervention to a commercial program. *Obesity* **15**, 155–164. (doi:10.1038/oby.2007.520)

10. Saperstein SL, Atkinson NL, Gold RS. 2007 The impact of Internet use for weight loss. *Obes. Rev.* **8**, 459–465. (doi:10.1111/j.1467-789X.2007.00374.x)

11. Harvey-Berino J, Pintauro S, Buzzell P, DiGiulio M, Gold BC, Moldovan C, Ramirez E. 2002 Does using the Internet facilitate the maintenance of weight loss? *Int. J. Obes.* **26**, 1254–1260. (doi:10.1038/sj.ijo.0802051)

12. Bahr DB, Browning RC, Wyatt HR, Hill JO. 2009 Exploiting social networks to mitigate the obesity epidemic. *Obesity* **17**, 723–728. (doi:10.1038/oby.2008.615)

13. Krukowski RA, Harvey-Berino J, Ashikaga T, Thomas CS, Micci N. 2008 Internet-based weight control: the relationship between web features and weight loss. *Telemed. e-Health* **14**, 775–782. (doi:10.1089/tmj.2007.0132)

14. Aral S, Walker D. 2011 Creating social contagion through viral product design: a randomized trial of peer influence in networks. *Manage. Sci.* **57**, 1623–1639. (doi:10.1287/mnsc.1110.1421)

15. Christakis NA, Fowler JH. 2012 Social contagion theory: examining dynamic social networks and human behavior. *Stat. Med.* **32**, 556–577. (doi:10.1002/sim.5408)

16. Cohen-Cole E, Fletcher JM. 2008 Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *J. Health Econ.* **27**, 1382–1387. (doi:10.1016/j.jhealeco.2008.04.005)

17. Christakis NA, Fowler JH. 2007 The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**, 370–379. (doi:10.1056/NEJMsa066082)

18. The Framingham Heart Study. 1948 See http://www.framinghamheartstudy.org/ and http://www.nhlbi.nih.gov/resources/obesity/pop-studies/framingham.htm.

19. Trogdon JG, Nonnemaker J, Pais J. 2008 Peer effects in adolescent overweight. *J. Health Econ.* **27**, 1388–1399. (doi:10.1016/j.jhealeco.2008.05.003)

20. Valente TW, Fujimoto K, Chou CP, Spruijt-Metz D. 2009 Adolescent affiliations and adiposity: a social network analysis of friendships and obesity. *J. Adolesc. Health* **45**, 1202–1204. (doi:10.1016/j.jadohealth.2009.01.007)

21. Leahey T, LaRose JG, Fava JL, Wing RR. 2011 Social influences are associated with BMI and weight loss intentions in young adults. *Obesity* **19**, 1157–1162. (doi:10.1038/oby.2010.301)

22. Wing RR, Jeffery RW. 1999 Benefits of recruiting participants with friends and increasing social support for weight loss and maintenance. *J. Consult. Clin. Psychol.* **67**, 132–138. (doi:10.1037/0022-006X.67.1.132)

23. Hawe P, Shiell A, Riley T. 2009 Theorising interventions as events in systems. *Am. J. Commun. Psychol.* **43**, 267–276. (doi:10.1007/s10464-009-9229-9)

24. Centola D. 2011 An experimental study of homophily in the adoption of health behavior. *Science* **334**, 1269–1272. (doi:10.1126/science.1207055)

25. Heckman JJ. 1976 The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Ann. Econ. Soc. Meas.* **5**, 475–492.

26. Perri MG, Nezu AM, Patti ET, McCann KL. 1989 Effect of length of treatment on weight loss. *J. Consult. Clin. Psychol.* **57**, 450–452. (doi:10.1037/0022-006X.57.3.450)

27. Wadden TA, Crerand CE, Brock J. 2005 Behavioral treatment of obesity. *Psychiatr. Clin. North Am.* **28**, 151–170, ix. (doi:10.1016/j.psc.2004.09.008)

28. Williamson DA, Anton S, Han H. 2010 Early behavioral adherence predicts short and long-term weight loss in the POUNDS LOST study. *J. Behav. Med.* **33**, 305–314. (doi:10.1007/s10865-010-9253-0)

29. Acharya SD, Elci OU, Sereika SM, Music E, Styn MA, Turk MW, Burke LE. 2009 Adherence to a behavioral weight loss treatment program enhances weight loss and improvements in biomarkers. *Patient Prefer Adherence* **3**, 151–160.

30. Aarts H, Gollwitzer PM, Hassin RR. 2004 Goal contagion: perceiving is for pursuing. *J. Pers. Soc. Psychol.* **87**, 23–27. (doi:10.1037/0022-3514.87.1.23)

31. Deci EL, Friedman R, Elliot AJ, Moller AC, Aarts H. 2010 Motivational synchronicity: priming motivational states via exposure to others. *Motiv. Emotion* **34**, 34–38. (doi:10.1007/s11031-009-9151-3)

32. Costa LDF. 2007 Characterization of complex networks: a survey of measurements. *Adv. Phys.* **56**, 167–242. (doi:10.1080/00018730601170527)

33. Carmi S, Havlin S, Kirkpatrick S, Shavitt Y, Shir E. 2007 A model of Internet topology using *k*-shell decomposition. *Proc. Natl Acad. Sci. USA* **104**, 11 150–11 154. (doi:10.1073/pnas.0701175104)

34. Gallos LK, Barttfeld P, Havlin S, Sigman M, Makse HA. 2012 Collective behavior in the spatial spreading of obesity. *Sci. Rep.* **2**, 454. (doi:10.1038/srep00454)

35. Pei S, Muchnik L, Andrade JS, Zheng Z, Makse HA. 2014 Searching for superspreaders of information in real-world social media. *Sci. Rep.* **4**, 5547. (doi:10.1038/srep05547)

36. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA. 2010 Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893. (doi:10.1038/nphys1746)

37. Pei S, Makse HA. 2013 Spreading dynamics in complex networks. *J. Stat. Mech.* **2013**, P12002. (doi:10.1088/1742-5468/2013/12/P12002)

38. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwanga DU. 2006 Complex networks: structure and dynamics. *Phys. Rep.* **424**, 175–308. (doi:10.1016/j.physrep.2005.10.009)

39. Newman MEJ. 2003 The structure and function of complex networks. *SIAM Rev.* **45**, 167–256. (doi:10.1137/S003614450342480)

40. Pittel B, Spencer J, Wormald N. 1996 Sudden emergence of a giant k-core in a random graph. *J. Combin. Theory* **67**, 111–151. (doi:10.1006/jctb.1996.0036)