www.sciencemag.org/cgi/content/full/science.1215842/DC1

Supporting Online Material for

Identifying Influential and Susceptible Members of Social Networks

Sinan Aral* and Dylan Walker*

*To whom correspondence should be addressed. E-mail: sinan@stern.nyu.edu; dwalker@stern.nyu.edu

Published 21 June 2012 on *Science* Express
DOI: 10.1126/science.1215842

This PDF file includes:

Materials and Methods

Figs. S1 to S12

Tables S1 to S9

Supporting Online Materials for

Identifying Influential and Susceptible Members of Social Networks

Data and Descriptive Statistics

The experiment was conducted over a 44-day period during which 7730 users adopted the application and sent 41,686 automated notifications[1] to randomly chosen targets amongst their 1.3M distinct peers, resulting in 976 peer adoptions. The randomization took place at the level of the local ego network, meaning that messages were randomized across the peers of every adopting user such that each peer of an adopting user had the same likelihood of receiving a randomized automated notification. Tables S1-S3 display descriptive statistics for the number of notifications sent and received by application users and their peers, respectively, and the subsequent adoption response according to age, gender and relationship status.

Table S1. Descriptive Statistics of User and Peer Demographics

	Number of Users	Number of Peers
Age 0-18	458	63063
Age 18-23	343	65606
Age 23-31	439	62176
Age 31+	959	69100
Age Unreported	5531	1036257
Male	867	134866
Female	1867	172406
Gender Unreported	4996	988930
Single	513	65410
In a Relationship	255	39536
Engaged	70	9494
Married	485	33561
Complicated	38	4775
Relationship Unreported	6369	1143426

Notes: The table reports the descriptive statistics concerning the demographic distributions of user and peer attributes for gender, age, and relationship status.

[1] The number of notifications shown here excludes any notifications that were sent from an application user to his peer at any time after the peer adopted.

Table S2. Descriptive Statistics of Peer Adoption Response in Local Networks of Users	Number of Notifications Sent	Average Number of Adopters in Local Network	Average Number of Adopters per Notification Sent
Age 0-18	2581	0.1659	6.43e-5
Age 18-23	1339	0.0875	6.53e-5
Age 23-31	1381	0.0661	4.79e-5
Age 31+	3486	0.0885	2.54e-5
Male	3005	0.0853	2.83e-5
Female	8700	0.1050	1.21e-5
Single	2805	0.1520	5.42e-5
In a Relationship	1551	0.1176	7.58e-5
Engaged	667	0.1143	1.71e-4
Married	2481	0.1052	4.24e-5
Complicated	430	0.1842	4.28e-4

Notes: The table reports the descriptive statistics concerning number of notifications sent by application users and the peer adoption response in the local networks of users according to user's gender, age, and reported relationship status

Table S3. Descriptive Statistics of Peer Adoption	Number of Notifications Received	Number of Peers Who Adopted	Average Number of Adopting Peers per Notification Received
Age 0-18	2641	91	3.45e-2
Age 18-23	2534	69	2.72e-2
Age 23-31	2388	43	1.80e-2
Age 31+	3619	117	3.23e-2
Male	6065	140	2.31e-2
Female	8422	267	3.17e-2
Single	1797	96	5.34e-2
In a Relationship	1243	40	3.22e-2
Engaged	303	9	2.97e-2
Married	1086	56	5.16e-2
Complicated	153	4	2.61e-2

Notes: The table reports the descriptive statistics concerning number of notifications received by peers and the resulting response according to peer's gender, age, and reported relationship status

Table S1 reports demographic distributions of user and peer attributes for gender, age, and relationship status. The first column of Tables S2 and S3 report the number of notifications sent by users to their local network peers and the number of notifications received by peers according to age, gender and relationship status attributes. The number of notifications sent by a user to his peers is a function of their application activity and limitations on the maximum number of notifications sent set by Facebook policy. An examination of these statistics reveals that female application users sent more than 2.5 times as many notifications as males. Users that reported their relationship status as *"Single"* sent the most notifications, followed by "*Married*," "*In a Relationship*," "*Engaged*," "*It's Complicated*," in descending order. While recipient targets of notifications are randomized at the ego network level, the number of notifications received by a peer is a function of the application activity of the peer's adopter friend (the application user). Although each peer of an application user has the same expected probability of receiving a notification, the number of notifications received by peers of an application user may depend on correlations between the application user's attributes and the attributes of their peers. For example, male users may tend to have more female peers (a heterophilous structure) making women more likely to receive notifications from men on aggregate. As Table S2 column 1 indicates, female peers received on average 130% more notifications than male peers. Peers that reported their relationship status as "*Single*" received the most notifications, followed by "*In a Relationship*," "*Married*," "*Engaged*," and "*It's Complicated*" in descending order. Our randomization procedure and subsequent analysis control for such systematic correlations by randomly distributing notifications to target peers of the same application user and controlling for the number of notifications received by peers.

Recruitment Campaign

At the beginning of the experiment, we designed and ran an advertising campaign in collaboration with a Facebook advertising firm to recruit a representative population of Facebook users to the experimental application. The advertising firm was instructed to display advertisements such that the likelihood that the recruited population was a representative sample of the Facebook population was maximized. Advertisements were subsequently displayed to users through advertising space within Facebook and within existing Facebook applications in such a way as to maximize the likelihood that representative proportions of the demographic characteristics of Facebook users were captured. The recruitment campaign cost a total of $6000 to recruit 7730 usable experimental subjects, or 78 cents per recruit. The campaign was conducted in three waves throughout the duration of the experiment to recruit a population of experimental subjects that consisted of 7730 application users and 1.3M distinct peers. Of the 8,910 advertising related installations of the application, 7730 users continued to fully install and use the application sufficiently to grant permission for the application to send notifications on their behalf.

The application was also publically listed in Facebook's Application Directory and so was available to anyone on Facebook. Details of the campaign are displayed in Table S4.

Table S4. Recruitment Statistics Describing the Initial Advertising Campaign

Wave	**Impressions**	**Clicks**	**Advertising Related Installations**	**Installations**
1 (Day 0)	18,264,600	12,334	3,072	3,714
2 (Day 15)	20,912,880	25,709	2,619	3,474
3 (Day 20)	19,957,640	7,624	3,219	4,039
Total	59,135,120	45,667	8,910	11,227

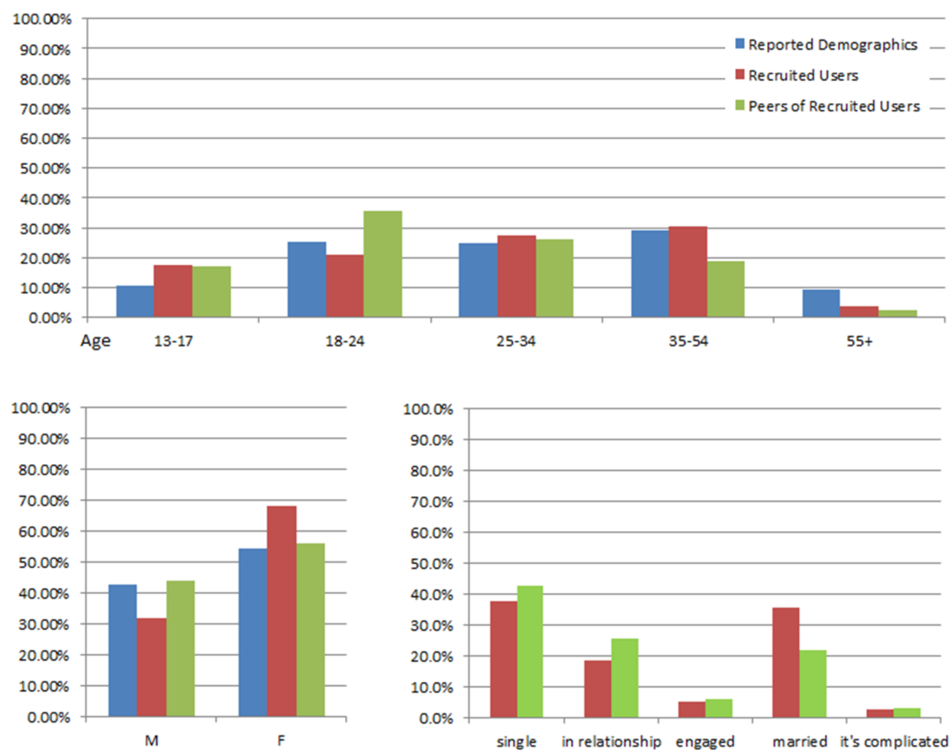
Representativeness of Sample

While we took the steps outlined above to ensure that application users and their peers were as representative of the Facebook population as possible, our analysis and influence estimates do not depend upon recruiting a fully representative sample. While deviations of the demographics of application users and their peers from the larger population may introduce more variance (and thus wider confidence intervals) in estimates of influence, susceptibility to influence and spontaneous adoption hazards for underrepresented demographic categories, estimates of the coefficients themselves are not subject to any systematic bias because randomization eliminates any selection effects. Nonetheless, we find that all demographic categories are well represented in our population of application users and their peers and compare this population to the best available data on Facebook population demographics to test the representativeness of our sample to the larger Facebook population.

Facebook does not publish or make available any official data regarding the demographics of its user population, however, we compared basic demographics of age and gender to a recent report published online by istrategylabs.com, a social targeting advertisement service. Figure S1 shows a comparison of the demographics of the recruited user population as well as of peers of recruited users to the published demographics. We find that the demographics of users in our study were generally representative of the Facebook population at the time the study was conducted and the published

Facebook demographics fall within one standard deviation of study population sample means.[2] Peers of recruited users are also well represented across demographic categories, though the peer population sample has more individuals in the 18-24 age range, less individuals in the 35-54 age range, and is more representative of the broader population in terms of the gender distribution than the population of recruited users.

Figure S1



Experimental Design

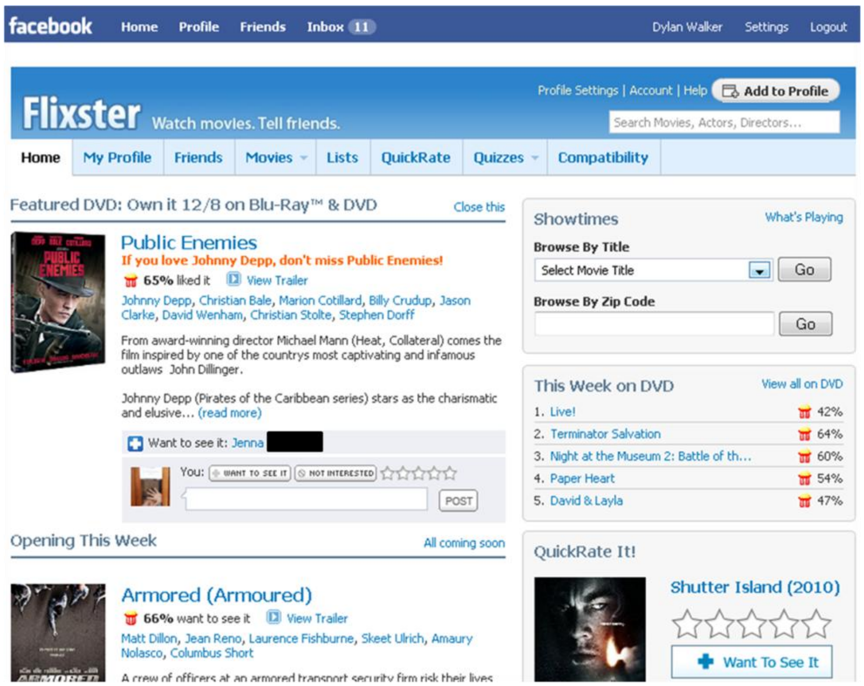
Figure S2 displays the front page of an example Facebook application that is similar to the one that we studied and is representative of the view a user sees when using the application.[3] Figure S3

[2] It is important to note that data published about the demographics of the Facebook population from services such as istrategylabs.com may reflect registered users (rather than, for example, active users) and are themselves statistically sampled. Thus, even in the case where a recruited study population differs from sample statistics published online, it may actually be more representative of active Facebook users than published statistics which may also count registered but inactive users.

[3] In accordance with the wishes of the Facebook application developer with which we partnered, the Facebook application displayed here is not the application used for the study. However, it is a similar Facebook application related to the movie industry that provides nearly identical functionality. The pictures displayed here are solely for the purpose of illustration.

displays the notification inbox, where a Facebook user may view and click on notifications delivered to her inbox. The notification inbox is private and only visible to users logged into their Facebook accounts. It is not visible to peers visiting other users' profile pages.

Figures S2 (left) and S3 (right)

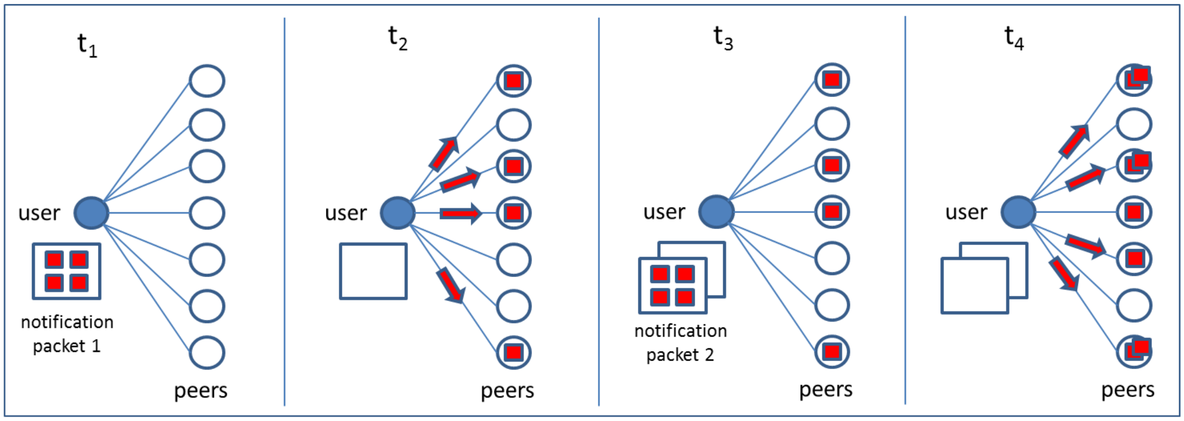


The procedure to randomize the delivery targets of automated notifications is illustrated in Figure S4. As application users engaged in actions on the application during the course of normal use, for example when they rated a movie or friended a celebrity, packets of notifications informing their friends of their use of the application were automatically generated in response to those actions and delivered to their randomly targeted Facebook peers. Each packet contained a fixed number of notifications, each of which was randomly targeted to a specific peer of the application user. This process was repeated for each action the user took on the application.[4] The number of notifications that a particular peer of an application user received at any given time was a function of a random Poisson process that depended only on the

[4] Facebook enforces a maximum limit of the number of notifications that an application can send on behalf of its users, as a spam prevention measure.

application user's sending rate (or the total number of notifications sent) and their network degree (the number of social network peers).

Figure S4



At time t_1, a packet of notifications (notification packet 1) was generated. At time t_2, peer targets were chosen randomly to be message recipients and were sent notifications from notification packet 1. At time t_3, a second packet of notifications was generated (notification packet 2). At time t_4, another set of peer targets were chosen randomly to be message recipients and were sent notifications from notification packet 2. Importantly, this second set of randomly chosen peer targets was selected independently of the set of peers randomly chosen to receive messages from the first notification packet. As a result, at any time *t*, a peer could have received zero, one, two, or more notifications from the application user. We define the quantity of influence-mediating notifications received by any particular peer *j* as $\mathrm{N_j(t)}$. This quantity, the number of notifications received by peer *j* at time t, is the randomized treatment (rather than an observed proxy for the treatment). It reflects the peer's "risk group," the extent to which they have been exposed to influence-mediating messages from their friend. Randomized treatment of peers occurred dynamically throughout the course of the experiment and was codified by the dynamic treatment variable $\mathrm{N_j(t)}$. To handle dynamic changes in randomized treatment in our hazard model estimation, we employed interval censoring. When any peer received a notification at time *t*, they were censored out of their prior risk group, $\mathrm{N_j(t-\epsilon)}$ (where ϵ is some infinitesimal time), and censored into their new risk group,

$N_j(t+\epsilon) = N_j(t-\epsilon) + 1$. This censoring procedure correctly parameterizes our ignorance of what might have happened had the peer not received an additional notification at time *t*.

Integrity of Randomization

To assure the integrity of our randomization procedure, we evaluated conditional logistic regression models estimating the number of notifications received by peers as a function of peer age, gender, and relationship status as well as the number of common friends between the peer and her application user friend (a measure of the embeddedness of the relationship and a proxy for the strength of the tie). Conditional logistic regression models are appropriate as they evaluate the dependence of the number of notifications received on peer attributes, conditional on the stratified grouping of peers with their common application user friend whose own activity on the application determines the rate at which peers receive notifications and the total number of notifications sent to all peers. The results, shown in Table S5 reveal no statistically significant dependence of the number of notifications received on any of the peer attributes considered, confirming the integrity of our randomization procedure.

Table S5: Integrity of Randomization via Conditional Logistic Regression Models

	β	$exp(\beta)$	$se(\beta)$	**z**	**P-value**
Number common friends	7.41E-05	1.000	0.000	0.228	0.820
Age 0-18	-5.08E-03	0.995	0.027	-0.190	0.850
Age 18-23	-1.54E-02	0.985	0.027	-0.578	0.560
Age 23-31	1.75E-03	1.002	0.027	0.065	0.950
Age 31+	6.12E-03	1.006	0.024	0.260	0.790
Male	2.12E-02	1.021	0.021	1.002	0.320
Female	1.28E-02	1.013	0.019	0.660	0.510
Single	-1.15E-03	0.999	0.029	-0.040	0.970
In Relationship	4.01E-02	1.041	0.034	1.187	0.240
Engaged	-7.17E-02	0.931	0.063	-1.134	0.260
Married	2.34E-02	1.024	0.036	0.650	0.520
It’s Complicated	9.93E-02	1.104	0.090	1.110	0.270

Notes: This table reports parameter estimates, standard errors, hazard ratios, z-scores, and p-values for the conditional logistic regression of a peer receiving one or more notifications conditional on her particular application user friend. The dependent variables indicate the peer’s attributes. The number of common friends is the number of friends a peer shares in common with her application user friend.

Statistical Results of Parameter Estimates Displayed in Forest Plot Figures 1-3.

Parameter estimates, confidence intervals and p-values for the forest plots described in Figures 1-3 in the paper are displayed in Tables S6 and S7. For example, our parameter estimates indicate that all else equal, the marginal effect of receiving an additional notification increases the hazard rate of adoption by 474% on average. In the Influence and Susceptibility Cox Proportional Hazards Model, the baseline represents individuals who do not report *age*, *gender*, and *relationship status* as part of their profile. In the Dyadic Cox Proportional Hazards Model, the baseline represents dyads in which the attributes are undefined or not reported for one or both members of the dyad (the individual and their peer).

Table S6: Estimates from Influence and Susceptibility Cox Proportional Hazards Model

	$\boldsymbol{\beta}$	**exp** $(\boldsymbol{\beta})$	$\boldsymbol{se(\beta)}$	**z**	**Pr(>|z|)**	**CI Lower .95**	**CI Upper .95**	**Robust** $\boldsymbol{se(\beta)}$	**Robust z**	**Robust Pr(>|z|)**	**Robust CI Lower .95**	**Robust CI Upper .95**
Treatment $(\boldsymbol{\beta_N})$												
# Notifications	1.747	5.736	0.045	38.543	< 2e-16	5.249	6.269	0.084	20.85	< 2e-16	4.868	6.760
Spontaneous Adoption of i $(\boldsymbol{\beta^i_{Spont}})$												
Age (0-18)	0.338	1.403	0.165	2.046	0.041	1.014	1.940	0.184	1.838	0.066	0.978	2.012
Age (18-23)	-0.389	0.678	0.234	-1.665	0.096	0.429	1.072	0.244	-1.597	0.110	0.421	1.092
Age (23-31)	-0.184	0.832	0.225	-0.816	0.415	0.535	1.294	0.268	-0.685	0.493	0.492	1.408
Age (>31)	-0.038	0.963	0.160	-0.237	0.813	0.704	1.316	0.169	-0.224	0.823	0.691	1.341
Male	-0.085	0.919	0.172	-0.495	0.620	0.656	1.286	0.191	-0.444	0.657	0.631	1.337
Female	0.072	1.075	0.132	0.545	0.586	0.830	1.392	0.150	0.478	0.633	0.800	1.443
Single	-0.129	0.879	0.151	-0.852	0.394	0.654	1.182	0.165	-0.777	0.437	0.636	1.216
Relationship	-0.185	0.831	0.210	-0.879	0.379	0.550	1.256	0.262	-0.706	0.480	0.497	1.389
Engaged	-0.330	0.719	0.414	-0.797	0.426	0.319	1.619	0.444	-0.743	0.457	0.301	1.716
Married	-0.326	0.722	0.186	-1.756	0.079	0.502	1.039	0.190	-1.720	0.085	0.498	1.047
Its Complicated	-0.125	0.883	0.419	-0.298	0.766	0.388	2.008	0.453	-0.275	0.783	0.363	2.146
Spontaneous Adoption of j $(\boldsymbol{\beta^j_{Spont}})$												
Age (0-18)	0.105	1.111	0.151	0.695	0.487	0.826	1.493	0.139	0.753	0.452	0.845	1.459
Age (18-23)	-0.028	0.972	0.160	-0.177	0.860	0.710	1.331	0.155	-0.183	0.855	0.718	1.317
Age (23-31)	-0.447	0.640	0.190	-2.353	0.019	0.441	0.928	0.181	-2.468	0.014	0.448	0.912
Age (>31)	0.433	1.542	0.136	3.176	0.001	1.181	2.015	0.133	3.264	0.001	1.189	2.001
Male	0.466	1.593	0.132	3.518	0.000	1.229	2.064	0.128	3.640	0.000	1.240	2.047
Female	0.894	2.444	0.112	7.957	0.000	1.961	3.046	0.111	8.020	0.000	1.965	3.041
Single	0.266	1.305	0.133	1.994	0.046	1.005	1.695	0.137	1.936	0.053	0.997	1.708
Relationship	-0.107	0.899	0.189	-0.567	0.571	0.621	1.301	0.187	-0.573	0.567	0.623	1.296
Engaged	-0.381	0.683	0.411	-0.926	0.354	0.305	1.529	0.362	-1.053	0.292	0.336	1.389
Married	0.310	1.363	0.162	1.911	0.056	0.992	1.873	0.165	1.881	0.060	0.987	1.883

Its Complicated	-0.633	0.531	0.641	-0.987	0.324	0.151	1.866	0.550	-1.151	0.250	0.181	1.560
Influence $(\boldsymbol{\beta_{Infl}})$												
Age (0-18)	-0.245	0.782	0.132	-1.853	0.064	0.604	1.014	0.146	-1.677	0.094	0.587	1.042
Age (18-23)	0.139	1.149	0.154	0.904	0.366	0.850	1.553	0.161	0.861	0.389	0.837	1.577
Age (23-31)	-0.125	0.882	0.238	-0.528	0.598	0.554	1.405	0.290	-0.433	0.665	0.500	1.557
Age (>31)	0.167	1.182	0.154	1.081	0.280	0.873	1.599	0.186	0.897	0.370	0.821	1.701
Male	0.154	1.166	0.140	1.101	0.271	0.887	1.534	0.161	0.955	0.340	0.850	1.600
Female	-0.243	0.784	0.102	-2.391	0.017	0.642	0.957	0.125	-1.942	0.052	0.613	1.002
Single	0.538	1.712	0.139	3.863	0.000	1.303	2.249	0.185	2.915	0.004	1.193	2.458
Relationship	-0.217	0.805	0.292	-0.743	0.457	0.454	1.426	0.282	-0.770	0.441	0.464	1.398
Engaged	0.115	1.121	0.345	0.332	0.740	0.570	2.207	0.306	0.375	0.708	0.616	2.043
Married	0.660	1.935	0.163	4.041	0.000	1.405	2.666	0.146	4.515	0.000	1.453	2.578
Its Complicated	-0.286	0.751	0.411	-0.695	0.487	0.336	1.682	0.293	-0.976	0.329	0.424	1.334
Susceptibility $(\boldsymbol{\beta_{Susc}})$												
Age (0-18)	0.072	1.074	0.109	0.660	0.510	0.868	1.330	0.102	0.704	0.482	0.880	1.312
Age (18-23)	-0.157	0.854	0.120	-1.306	0.192	0.675	1.082	0.107	-1.468	0.142	0.693	1.054
Age (23-31)	-0.110	0.895	0.130	-0.849	0.396	0.694	1.156	0.084	-1.322	0.186	0.760	1.055
Age (>31)	-0.192	0.825	0.112	-1.710	0.087	0.662	1.029	0.085	-2.261	0.024	0.698	0.975
Male	-0.259	0.772	0.091	-2.843	0.004	0.646	0.923	0.066	-3.918	0.000	0.678	0.879
Female	-0.388	0.678	0.071	-5.463	0.000	0.590	0.780	0.064	-6.024	0.000	0.598	0.770
Single	0.347	1.415	0.113	3.071	0.002	1.134	1.765	0.099	3.494	0.000	1.165	1.719
Relationship	0.349	1.417	0.171	2.036	0.042	1.013	1.983	0.152	2.293	0.022	1.052	1.910
Engaged	0.774	2.168	0.262	2.952	0.003	1.297	3.623	0.209	3.700	0.000	1.439	3.265
Married	0.014	1.014	0.147	0.094	0.925	0.759	1.354	0.135	0.102	0.919	0.778	1.322
Its Complicated	0.748	2.112	0.405	1.846	0.065	0.955	4.672	0.308	2.432	0.015	1.156	3.859

Notes: This table reports parameter estimates, hazard ratios, z-scores, confidence intervals and P-values for the Influence and Susceptibility Cox proportional hazards model that estimate the impact of a user's age, gender or relationship status on his hazard to influence peers to adopt and on the hazard that his peers will spontaneously adopt. The table summarizes the model of influenced and spontaneous adoption with age, gender and relationship status as independent variables, while controlling for the remaining attributes.

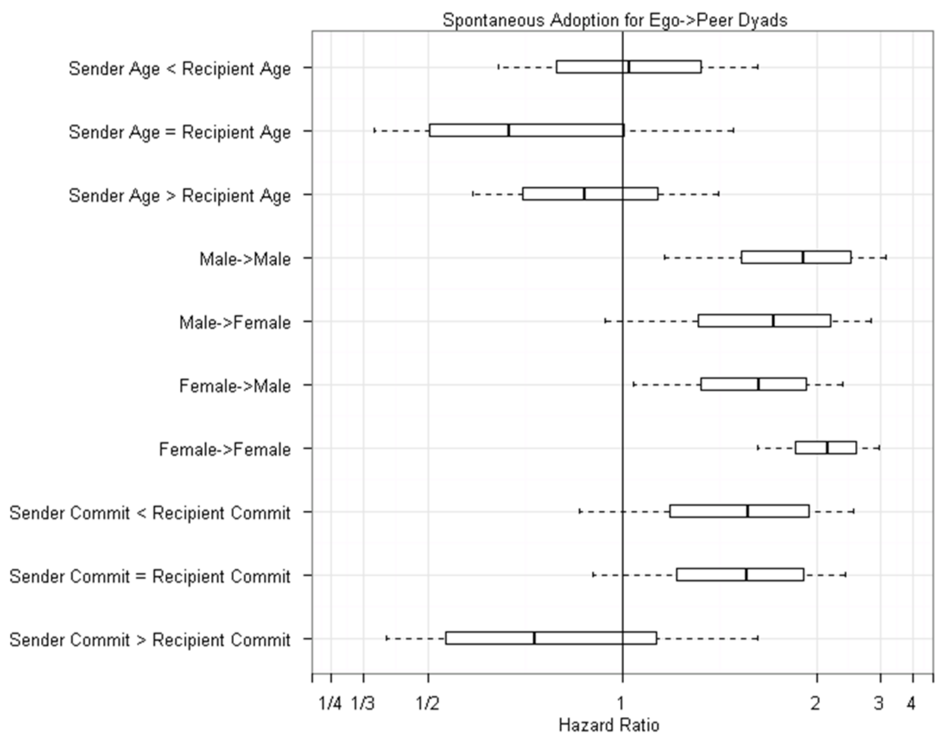
Table S7: Estimates from Dyadic Cox Proportional Hazards Model

	$\boldsymbol{\beta}$	**exp** ($\boldsymbol{\beta}$	$\boldsymbol{se(\beta)}$	**z**	**Pr(>|z|)**	**CI Lower .95**	**CI Upper .95**	**Robust** $\boldsymbol{se(\beta)}$	**Robust z**	**Robust Pr(>|z|)**	**Robust CI Lower .95**	**Robust CI Upper .95**
Treatment ($\boldsymbol{\beta_N}$)												
# Notifications	1.596	4.934	0.029	55.009	< 2e-16	0.062	25.945	< 2e-16	4.374	5.567	1.596	4.934
Spontaneous Adoption ($\boldsymbol{\beta^i_{Spont}}$)												
S Age < R Age	-0.102	0.903	0.201	-0.506	0.613	0.196	-0.518	0.604	0.615	1.327	-0.102	0.903
S Age = R Age	-0.343	0.710	0.377	-0.909	0.363	0.346	-0.990	0.322	0.360	1.399	-0.343	0.710
S Age > R Age	0.020	1.020	0.213	0.092	0.927	0.208	0.094	0.925	0.679	1.532	0.020	1.020
Male → Male	0.627	1.872	0.271	2.314	0.021	0.261	2.399	0.016	1.122	3.125	0.627	1.872
Male → Female	0.492	1.636	0.275	1.791	0.073	0.274	1.798	0.072	0.957	2.797	0.492	1.636
Female → Male	0.434	1.543	0.213	2.038	0.042	0.207	2.100	0.036	1.029	2.313	0.434	1.543
Female → Female	0.757	2.131	0.164	4.606	0.000	0.166	4.554	0.000	1.539	2.952	0.757	2.131
S Com < R Com	-0.257	0.773	0.348	-0.738	0.461	0.349	-0.736	0.462	0.390	1.534	-0.257	0.773
S Com = R Com	0.389	1.475	0.237	1.643	0.100	0.239	1.624	0.104	0.923	2.358	0.389	1.475
S Com > R Com	0.394	1.483	0.270	1.460	0.144	0.262	1.504	0.132	0.888	2.479	0.394	1.483
Influence ($\boldsymbol{\beta_{Infl}}$)												
S Age < R Age	0.323	1.381	0.161	2.012	0.044	0.160	2.017	0.044	1.009	1.890	0.323	1.381
S Age = R Age	0.676	1.965	0.324	2.082	0.037	0.215	3.144	0.002	1.290	2.995	0.676	1.965
S Age > R Age	0.105	1.111	0.167	0.629	0.529	0.113	0.929	0.353	0.890	1.386	0.105	1.111
Male → Male	-0.106	0.899	0.188	-0.563	0.573	0.193	-0.550	0.582	0.616	1.313	-0.106	0.899
Male → Female	-0.351	0.704	0.154	-2.284	0.022	0.185	-1.898	0.058	0.490	1.012	-0.351	0.704
Female → Male	0.033	1.034	0.184	0.182	0.855	0.164	0.204	0.838	0.750	1.426	0.033	1.034
Female → Female	-0.343	0.710	0.110	-3.119	0.002	0.146	-2.343	0.019	0.533	0.945	-0.343	0.710
S Com < R Com	0.697	2.009	0.349	1.997	0.046	0.290	2.401	0.016	1.137	3.549	0.697	2.009
S Com = R Com	0.533	1.704	0.253	2.111	0.035	0.241	2.211	0.027	1.062	2.734	0.533	1.704
S Com > R Com	-0.153	0.858	0.572	-0.268	0.789	0.445	-0.343	0.731	0.358	2.055	-0.153	0.858

Notes: This table reports parameter estimates, hazard ratios, confidence intervals and P-values for the Cox proportional hazard model that estimates the impact of dyadic attributes of a sender/(potential)-recipient pair on the hazard that the potential recipient in the dyad will adopt via influence and on the hazard that he will spontaneously adopt. Dyadic attributes considered include indicators of whether the Sender is older, younger or the same age as the recipient; the possible gender combinations of Sender and Recipient; and whether the Sender is in a relationship that is less, equally or more committed than the relationship the Recipient is in. The table summarizes the model of influenced and spontaneous adoption pertaining to age-related, gender-related and relationship status-related dyadic measures, while controlling for the remaining dyadic attributes.

The hazard ratios for spontaneous adoption estimates obtained from dyadic models indicate the hazard for an individual to have a particular peer (ego->peer dyad) spontaneously adopt in the absence of influence. They are displayed below in a forest plot for the purposes of completeness:

Figure S5



Goodness of Fit

We employed several tests to assess specification and goodness-of-fit of the influence and susceptibility proportional hazards model and the dyadic peer-to-peer influence proportional hazards model. Cox proportional hazard models employ iterative fitting procedures to obtain estimates that maximize pseudo log-likelihood. The pseudo log-likelihood of the intercept-only model as well as the pseudo log-likelihood of the model with all included dependent covariates, the Likelihood Ratio, Wald and Score Tests, as well as concordance probability assessments of these models are all reported in Table S8. The Likelihood Ratio (LRT) Test evaluates the likelihood of the data under the fitted model relative to the null (intercept only) model and the associated test statistic converges to a chi-squared distribution. The LRT test statistic for the influence and susceptibility model is 1470 over 45 degrees of freedom ($p < 1e-12$) indicating a significantly better fit for the full model. The Wald Test (WT) assesses the likelihood of the data under the fitted model in a manner similar to the LRT, but employs a Taylor series expansion around $\beta = \beta_{final}$ and adjusts for tied failure times. The Score Test (ST) assess the likelihood of the data under the fitted model in a manner similar to the WT, but employs a Taylor series expansion around $\beta = 0$, uses estimated clustered standard errors and adjusts for tied times. The LRT, WT, and ST test statistics for the influence and susceptibility model are LRT=1470, WT=2637, and ST=357.2 over 45 degrees of freedom ($p < 1e-12$) and for the dyadic peer-to-peer influence models are LRT=1274,WT=1271, and ST=272 over 23 degrees of freedom ($p < 1e-12$). These tests uniformly confirm a significantly better fit for the full model specifications over the null model specifications.

Table S8: Goodness of Fit Tests Influence and Susceptibility and Dyadic Peer-to-Peer Cox Proportional Hazards Models

	Log Likelihood (Intercept)	**Log Likelihood**	**DOF**	**Likelihood Ratio Test**	**Wald Test**	**Score Test**	**Concordance Probability**
Influence and Susceptibility	-13516.15	-12780.92	45	1470	2637	357.2	78%
Dyadic Peer-to-Peer	-13516.15	-12879.06	23	1274	1271	272	73%

To assess the extent to which survival times of peers were in accordance with their estimated hazards to fail (adopt), we employed concordance probability tests which compare the relative order of survival for all pairs of peers in the data to the expected relative order of survival under the fitted model. The concordance probability (the proportion of observed relative peer survivals that are in accordance with model predictions) associated with the influence and susceptibility model is 78%, indicating relative survival of peer pairs as compared to predicted relative survival occurs with reasonable probability. The concordance probability for the dyadic peer-to-peer is 73%, indicating that predicted relative survival order occurs with reasonable probability.

In addition to formal statistical tests of specification and goodness-of-fit, we performed graphical analysis of residuals for survival models. Plots of component + Martingale residuals vs. linear covariates assess the extent to which assumptions of covariate linearity hold. In our models, covariates are largely dichotomous, with the exception of *number of notifications received (nnr)*. Plots of component+Martingale residuals vs. *number of notifications received* are displayed in Table S9 for the influence and susceptibility and dyadic peer-to-peer influence models. These residuals indicate only a slight non-linearity (as evidenced) by the departure of the (solid) lowess curve from the (dotted line) linear fit. This departure occurs for *number of notifications received* driven by larger values (nnr>3). Since the bulk of peers (99%) received fewer notifications (nnr<3), it is unlikely that our model estimates are significantly impacted by this slight non-linearity displayed. Furthermore, because we focus on the modulating impact of dichotomous covariates on the response to receiving notifications and because peers with differing covariate values were equally likely to randomly receive any given number of notifications[5], the impact of any slight non-linearity on estimates of influence and susceptibility must be equal across peers with differing covariate values. Furthermore, the majority of comparison of influence and susceptibility are relative and so will not be affected by overall shifts of influence and susceptibility hazard estimates across all covariates.

Plots of scaled Schoenfeld residuals associated with model covariates across survival times assess the validity of the proportional hazards assumption. Linear trends in scaled Schoenfeld residuals associated with a particular covariate across survival times indicate that the proportional hazards assumption is violated for that covariate. Scaled Schoenfeld residual plots for the 45 model covariates in the influence and susceptibility model are displayed in Fig. S9, and for the dyadic peer-to-peer influence model in Fig. S11. We do not observe any significant trends, indicating the validity of the proportional hazards assumption.

[5] Our estimates of influence and susceptibility are robust to the inclusion of Frailty (to control for ego adopter identity / local network group effects) and controls for ego “dosage” and “interest” in the application, affirming that unobserved heterogeneity on the part of the sender (ego) does not affect the likelihood for peers with particular covariates to receive any given number of notifications.

Plots of dfbeta residuals across peer subject for model estimates assess the contribution of a given subject to the fitted estimation (β) (i.e., the relative change in the estimate when a given subject observation is omitted from the data). Plots of dfbeta residuals for the 45 covariates in the influence and susceptibility Cox proportional hazard model and the 23 covariates in the dyadic peer-to-peer influence Cox proportional hazard model are displayed in Figs. S10 and S12, respectively. These plots reveal that, overall, no single observation in the data exert a disproportionate impact on model estimates.

Robustness, Group Specific Heterogeneity and Non-Independence

Our analysis aggregates individual experiments that take place at the local ego network level. One potential concern in such circumstances is that peers of the same adopting user are not independent, but rather experience common group level shocks to their adoption likelihoods. Heterogeneity across local network neighborhoods can introduce bias if, for example, some adopters have more affinity for the product and send more messages than others, and if there is homophily in these preferences such that peers of high affinity adopters are more likely as a group to adopt the product than peers of other adopters. We took numerous steps to ensure that our results are not biased by group level heterogeneity.

First, we checked the robustness of our estimates to the most likely specific concerns regarding heterogeneity in observable characteristics and behaviors across adopting users. To test the robustness of our results to the concern that some adopters will send more notifications than others, we estimated the influence and susceptibility model controlling for the number of notifications sent by adopter *i* divided by *i*'s degree (which represents the number of notifications peers of *i* would expect to receive). This had no effect on any of the other parameters and was itself not significant. We also controlled for the adopter *i*'s degree and the number of notifications sent by adopter *i* separately. None of these specifications changed the results either. These results should dispel any concern that heterogeneity in the sending rate of *i* is affecting our results.

Second, we estimated alternative specifications as robustness checks. However, as we explain here, none of the alternative specifications are appropriate for our modeling aims. This discussion highlights the importance of matching model specification choices (and the subsequent interpretation of parameter estimates) to the specific scientific and policy making goals of the analysis. To account for group level heterogeneity and adopter specific effects, we fit an influence and susceptibility model that accounts for observable characteristics of the adopter and estimated a shared frailty (random group effects) specification to control for unobserved heterogeneity. The shared frailty specification models intragroup correlations by introducing an unobservable multiplicative effect α on the hazard, so that conditional on the frailty

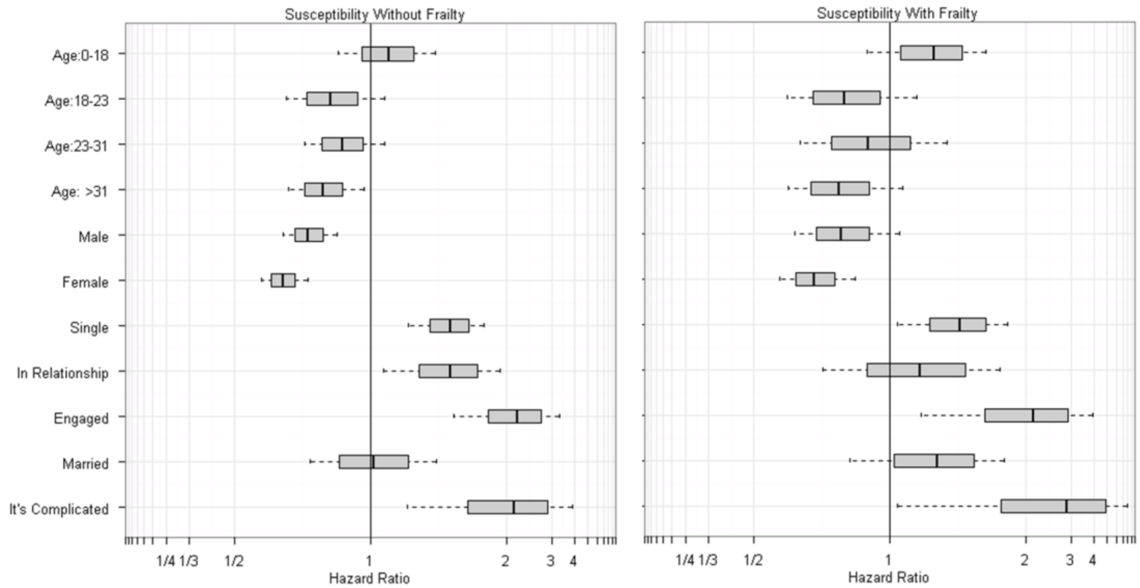
$$\lambda(t|\alpha) = \alpha_i \lambda(t)$$

where α_i is a random positive quantity with mean 1 and variance θ and i indexes the group – in this case the local ego network or the original adopter i. For any member of the ith group the hazard function is multiplied by the shared frailty α_i. Thus we estimated the influence and susceptibility model as follows:

$$\lambda\big(t, X_i, X_j, N_j|\alpha_i\big) = \alpha_i \lambda_0(t) \exp\big(N_j(t)\beta_N + X_i\beta_{Spont}^i + X_j\beta_{Spont}^j + N_j(t)X_i\beta_{Infl} + N_j(t)X_j\beta_{Susc}\big)$$

Results of the shared frailty model show that our susceptibility estimates are robust to the inclusion of random group effects (as well as to controls for adopters' observable characteristics and the inclusion of covariates for the number of notifications adopters send). The susceptibility estimates change somewhat but not substantially as shown below.

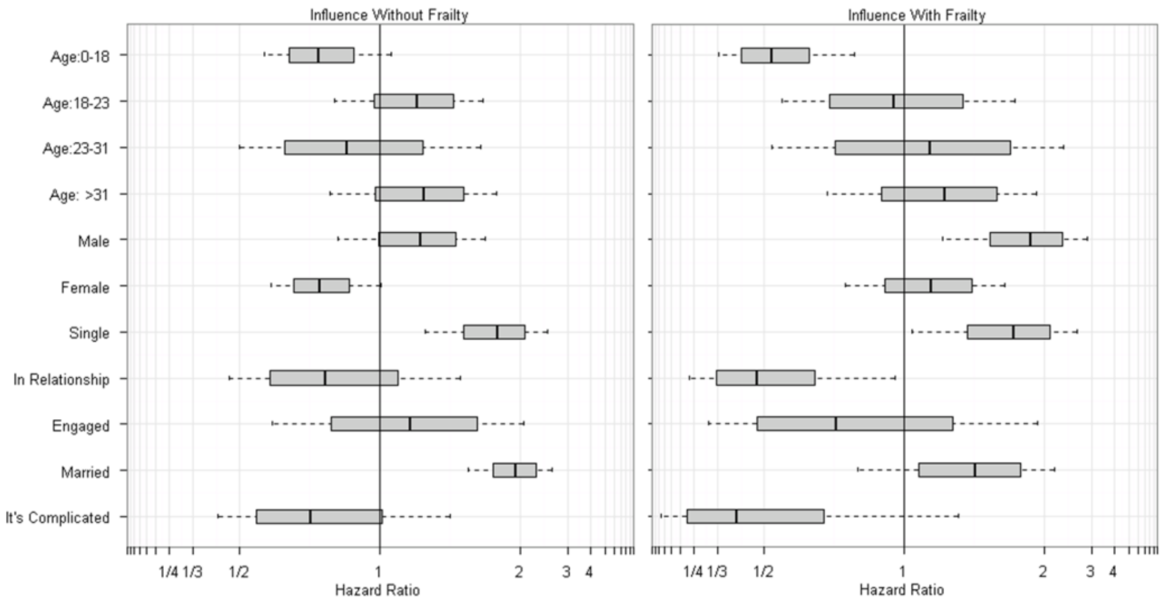
Figure S6



The influence terms change slightly more, but frailty specifications are not appropriate when estimating influence in our case because they model individual frailty with respect to the adopters (the message senders) (see Table S8 for full frailty results). They are not appropriate because we are not interested for example in estimating the effect of age on influence holding constant all unobservables – if experience is unobservable and creates influence, and if age and experience are correlated, we are less interested in estimating the effect of age net of experience, but rather whether age, for whatever reason, predicts influence. The reason we care about this effect rather than the effect of age net of all unobservables is that the policies we intend to inform with this analysis are not improved by understanding the causal effect of an additional year of age on influence, but rather by identifying characteristics of influential people whatever their underlying causes. This is because a government or firm policy targeting "influential" people would not attempt to exogenously change the age, gender or relationship status of a group of people in order to increase their influence, but would rather attempt to identify influential people in order to give them free products or anti-smoking education or some other intervention in the hopes of changing the behavior of their peers. The underlying causal relationship between individual characteristics and the magnitude of influence is not the key to optimizing this policy, but identifying correlates of influence is.

This is not to say that we are not interested in causal inference. We are interested in establishing the causal effect of peer influence on adoption (while controlling for example for the natural clustering of adoption amongst consumers with correlated preferences) and simultaneously estimating correlates of influence, rather than causes of influence, in other words, the characteristics of people who are more influential (e.g. men or women, the young or the old). The randomization procedure helps us establish causal influence controlling for the traditional confounds. The influence of an adopter on their peers via influence mediating messages is therefore better modeled by the inclusion of covariates for notifications and notifications moderated by user characteristics in the unified model. Dyadic models with and without frailty are shown below.

Figure S7



To account for the possibility that peers of the same adopters may not be i.i.d., we clustered the standard errors on the senders' local network. The significance of parameter estimates change only slightly and our results are robust to both clustering and shared frailty, indicating that variance introduced by within-network correlations in peer adoption do not significantly affect our findings. The results reported in the main text of the revision now use clustered standard errors.

Table S9: Estimates from Influence and Susceptibility Cox Proportional Hazards Model with Frailty						
	$\boldsymbol{\beta}$	**exp** $(\boldsymbol{\beta})$	$\boldsymbol{se}(\boldsymbol{\beta})$	**Pr(>|z|)**	**CI Lower .95**	**CI Upper .95**
Treatment $(\boldsymbol{\beta_N})$						
# Notifications	1.867	6.472	0.066	0.000	5.684	7.369
Spontaneous Adoption of i $(\boldsymbol{\beta^i_{Spont}})$						
Age (0-18)	0.338	1.403	0.165	0.041	1.014	1.940
Age (18-23)	-0.389	0.678	0.234	0.096	0.429	1.072
Age (23-31)	-0.184	0.832	0.225	0.415	0.535	1.294
Age (>31)	-0.038	0.963	0.160	0.813	0.704	1.316
Male	-0.085	0.919	0.172	0.620	0.656	1.286
Female	0.072	1.075	0.132	0.586	0.830	1.392
Single	-0.129	0.879	0.151	0.394	0.654	1.182
Relationship	-0.185	0.831	0.210	0.379	0.550	1.256
Engaged	-0.330	0.719	0.414	0.426	0.319	1.619
Married	-0.326	0.722	0.186	0.079	0.502	1.039
Its Complicated	-0.125	0.883	0.419	0.766	0.388	2.008
Spontaneous Adoption of j $(\boldsymbol{\beta^j_{Spont}})$						
Age (0-18)	0.105	1.111	0.151	0.487	0.826	1.493
Age (18-23)	-0.028	0.972	0.160	0.860	0.710	1.331
Age (23-31)	-0.447	0.640	0.190	0.019	0.441	0.928
Age (>31)	0.433	1.542	0.136	0.001	1.181	2.015
Male	0.466	1.593	0.132	0.000	1.229	2.064
Female	0.894	2.444	0.112	0.000	1.961	3.046
Single	0.266	1.305	0.133	0.046	1.005	1.695
Relationship	-0.107	0.899	0.189	0.571	0.621	1.301
Engaged	-0.381	0.683	0.411	0.354	0.305	1.529
Married	0.310	1.363	0.162	0.056	0.992	1.873
Its Complicated	-0.633	0.531	0.641	0.324	0.151	1.866
Influence $(\boldsymbol{\beta_{Infl}})$						
Age (0-18)	-0.245	0.782	0.132	0.064	0.604	1.014
Age (18-23)	0.139	1.149	0.154	0.366	0.850	1.553
Age (23-31)	-0.125	0.882	0.238	0.598	0.554	1.405
Age (>31)	0.167	1.182	0.154	0.280	0.873	1.599
Male	0.154	1.166	0.140	0.271	0.887	1.534
Female	-0.243	0.784	0.102	0.017	0.642	0.957
Single	0.538	1.712	0.139	0.000	1.303	2.249

Relationship	-0.217	0.805	0.292	0.457	0.454	1.426
Engaged	0.115	1.121	0.345	0.740	0.570	2.207
Married	0.660	1.935	0.163	0.000	1.405	2.666
Its Complicated	-0.286	0.751	0.411	0.487	0.336	1.682
Susceptibility ($\boldsymbol{\beta_{Susc}}$)						
Age (0-18)	0.072	1.074	0.109	0.510	0.868	1.330
Age (18-23)	-0.157	0.854	0.120	0.192	0.675	1.082
Age (23-31)	-0.110	0.895	0.130	0.396	0.694	1.156
Age (>31)	-0.192	0.825	0.112	0.087	0.662	1.029
Male	-0.259	0.772	0.091	0.004	0.646	0.923
Female	-0.388	0.678	0.071	0.000	0.590	0.780
Single	0.347	1.415	0.113	0.002	1.134	1.765
Relationship	0.349	1.417	0.171	0.042	1.013	1.983
Engaged	0.774	2.168	0.262	0.003	1.297	3.623
Married	0.014	1.014	0.147	0.925	0.759	1.354
Its Complicated	0.748	2.112	0.405	0.065	0.955	4.672

Methods for Contour & Network Plots

We calculated predicted influence and susceptibility scores for 12M Facebook users, based on their individual attributes, using the results from influence and susceptibility models. We define the predicted influence (susceptibility) score as the product of influence (susceptibility) hazard ratios for the attributes of age, gender and relationship status, as given by:

$$S_{Infl} = \prod_{a} \exp\left(\beta_{Infl,a}\right)$$

$$S_{Susc} = \prod_{a} \exp\left(\beta_{Susc,a}\right)$$

where $\beta_{Infl,a}$ ($\beta_{Susc,a}$) is the estimated influence (susceptibility) hazard associated with attribute a. For example, the predicted influence score for a 25 year old single male is given by:

$$S_{Infl} = \exp\left(\beta_{Infl,Age\ 23-31}\right) * \exp\left(\beta_{Infl,Male}\right) * \exp\left(\beta_{Infl,Single}\right)$$

This method of calculating predicted influence and susceptibility scores is consistent with the proportional hazards assumption implicit in the Cox models employed in our analysis.

The contour plots shown in Figure 4 are generated from predicted data using ridge regression surface modeling, a standard method for smoothing three-dimensional data. The method employs a regularizer proportional to the difference between first partial derivatives in neighboring bins, with the constant of proportionality chosen to be 2.5 to achieve sufficient smoothness. Contour plot 1 is generated from the set of unique values of predicted ego influence and ego susceptibility and the corresponding multiplicity for 12M individuals. Contour plots 2-4 are generated from the set of unique values of predicted ego influence (or susceptibility) and peer influence (or susceptibility) for 85M social relationships (edges) between the same 12M individuals.

Fig. S8: Component+Martingale Residuals for the Influence and Susceptibility and Dyadic Peer-to-Peer Influence Models

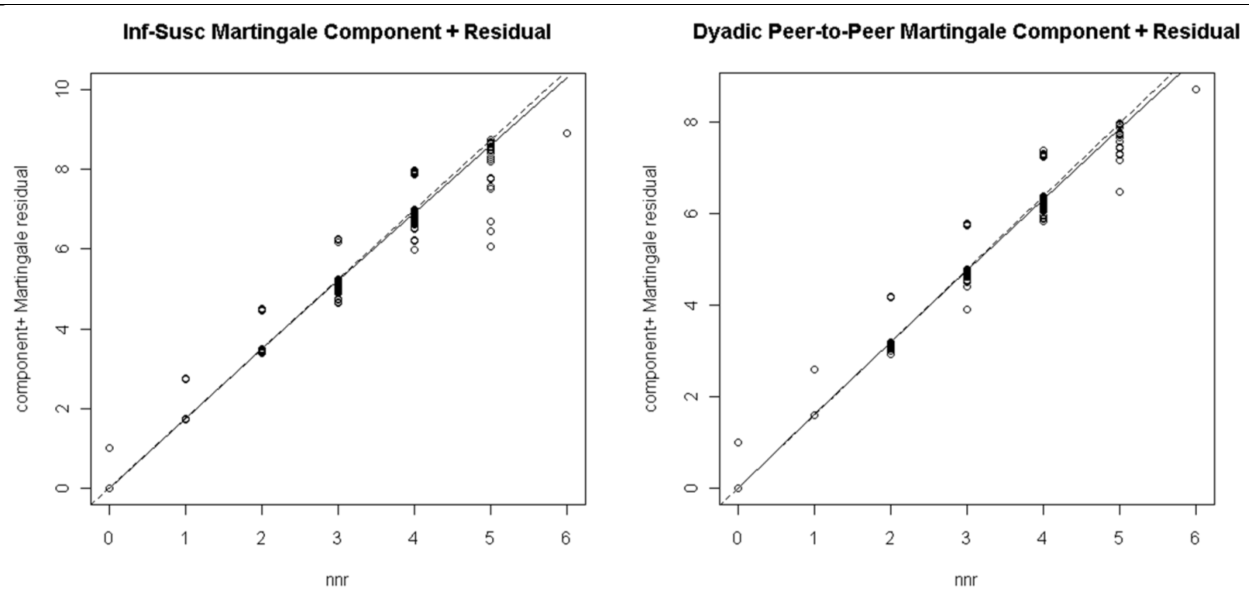
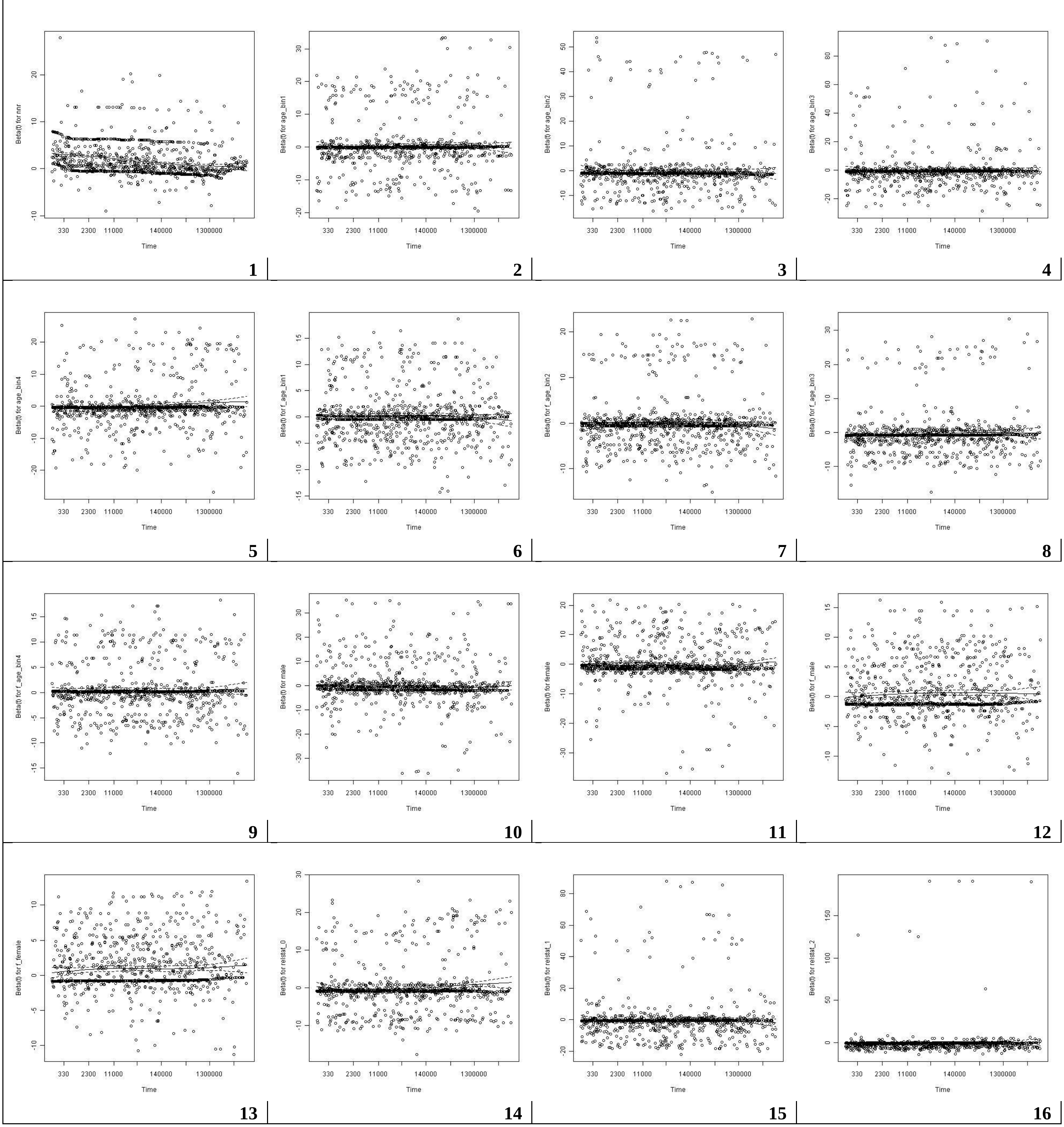
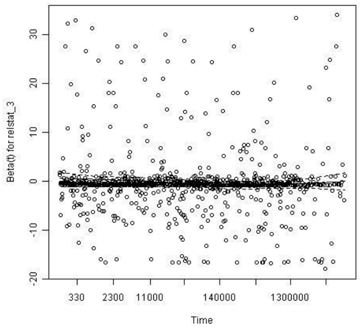
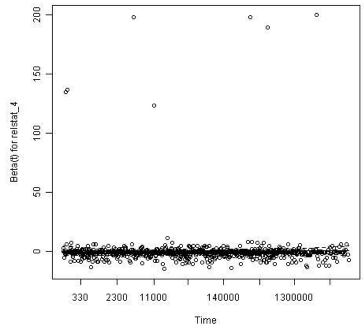


Fig. S9: Scaled Schoenfeld Residuals for the Influence and Susceptibility Cox Proportional Hazards Model

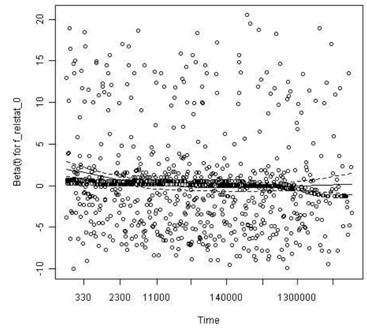




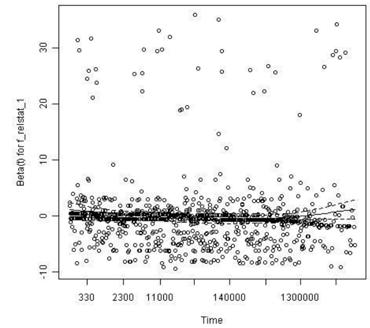
17



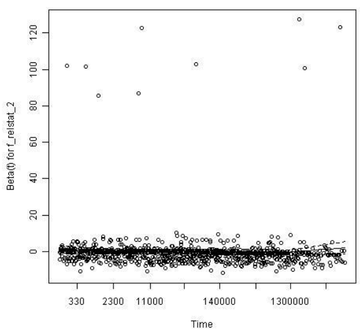
18



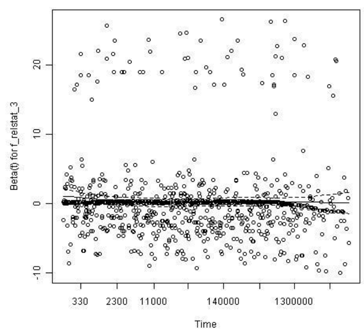
19



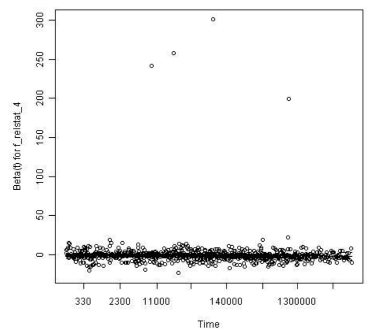
20



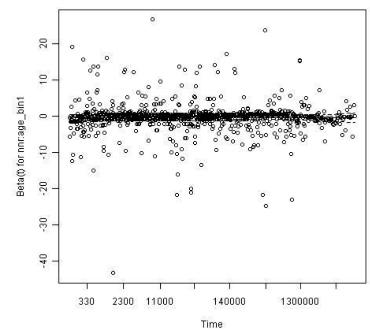
21



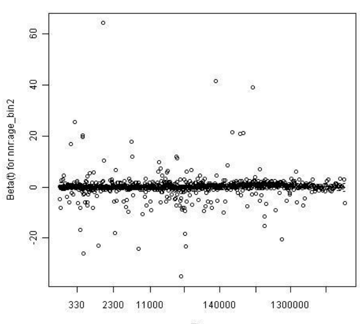
22



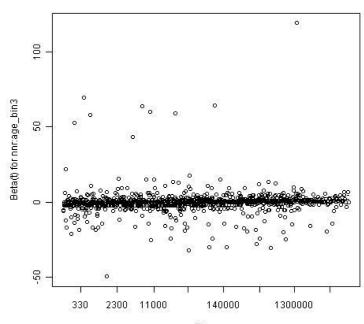
23



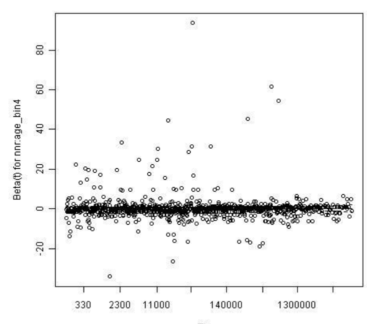
24



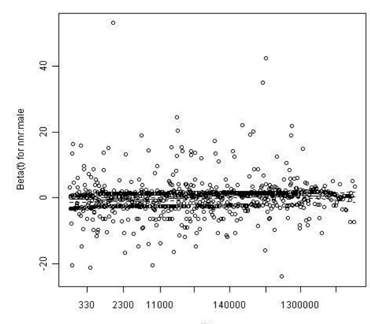
25



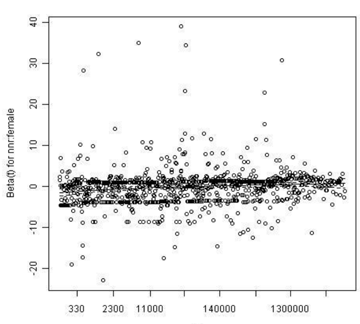
26



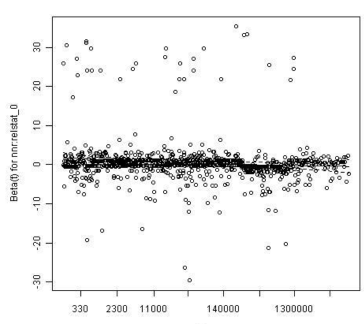
27



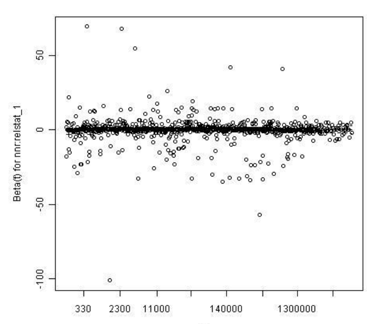
28



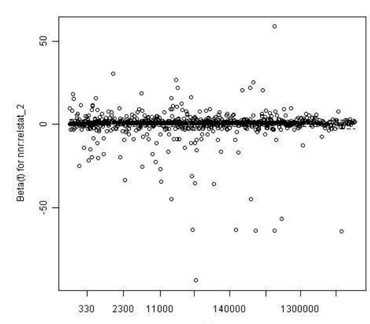
29



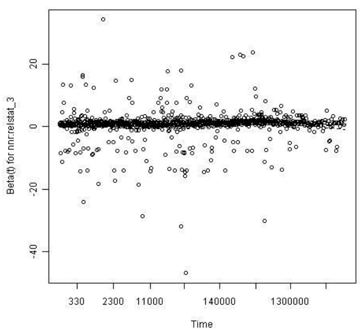
30



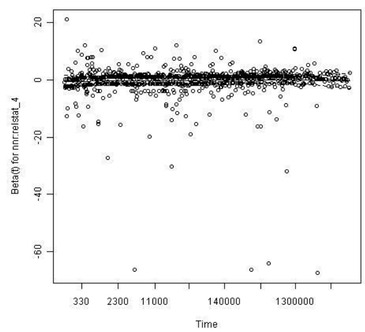
31



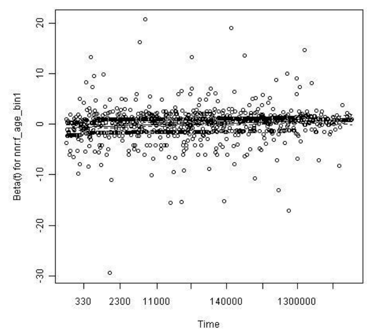
32



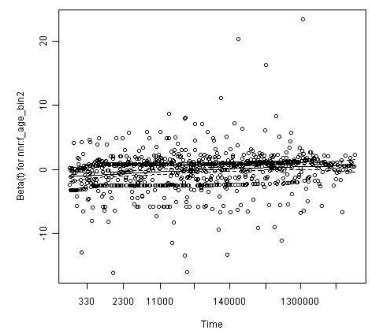
33



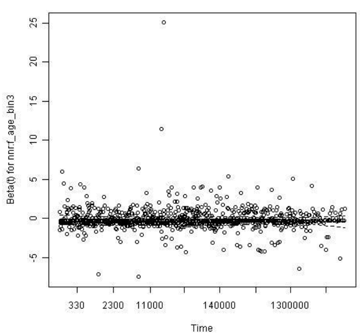
34



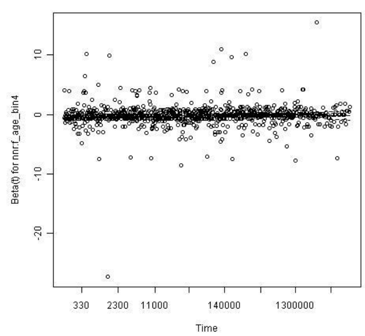
35



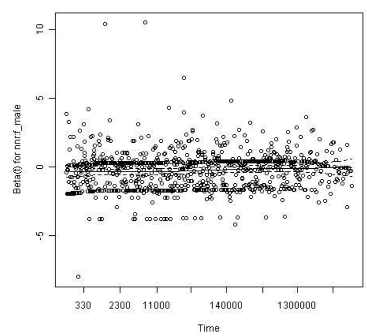
36



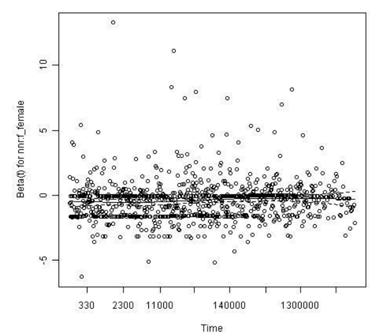
37



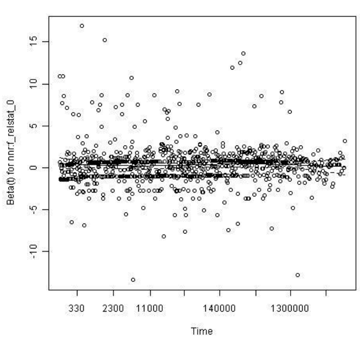
38



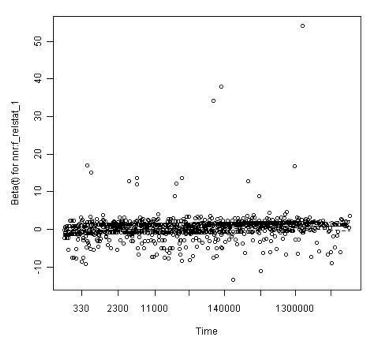
39



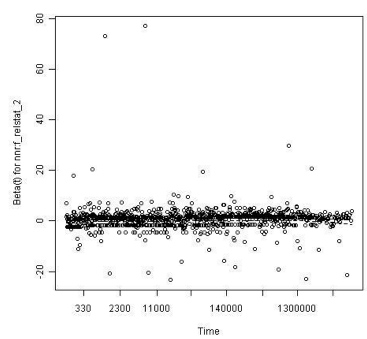
40



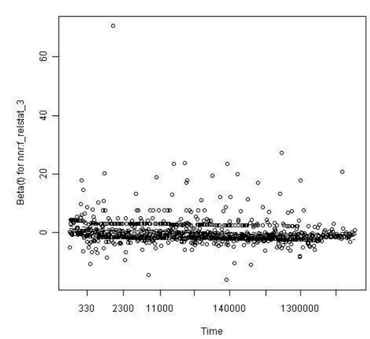
41



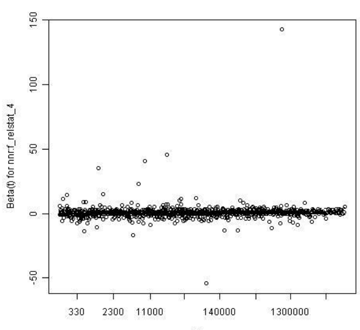
42



43

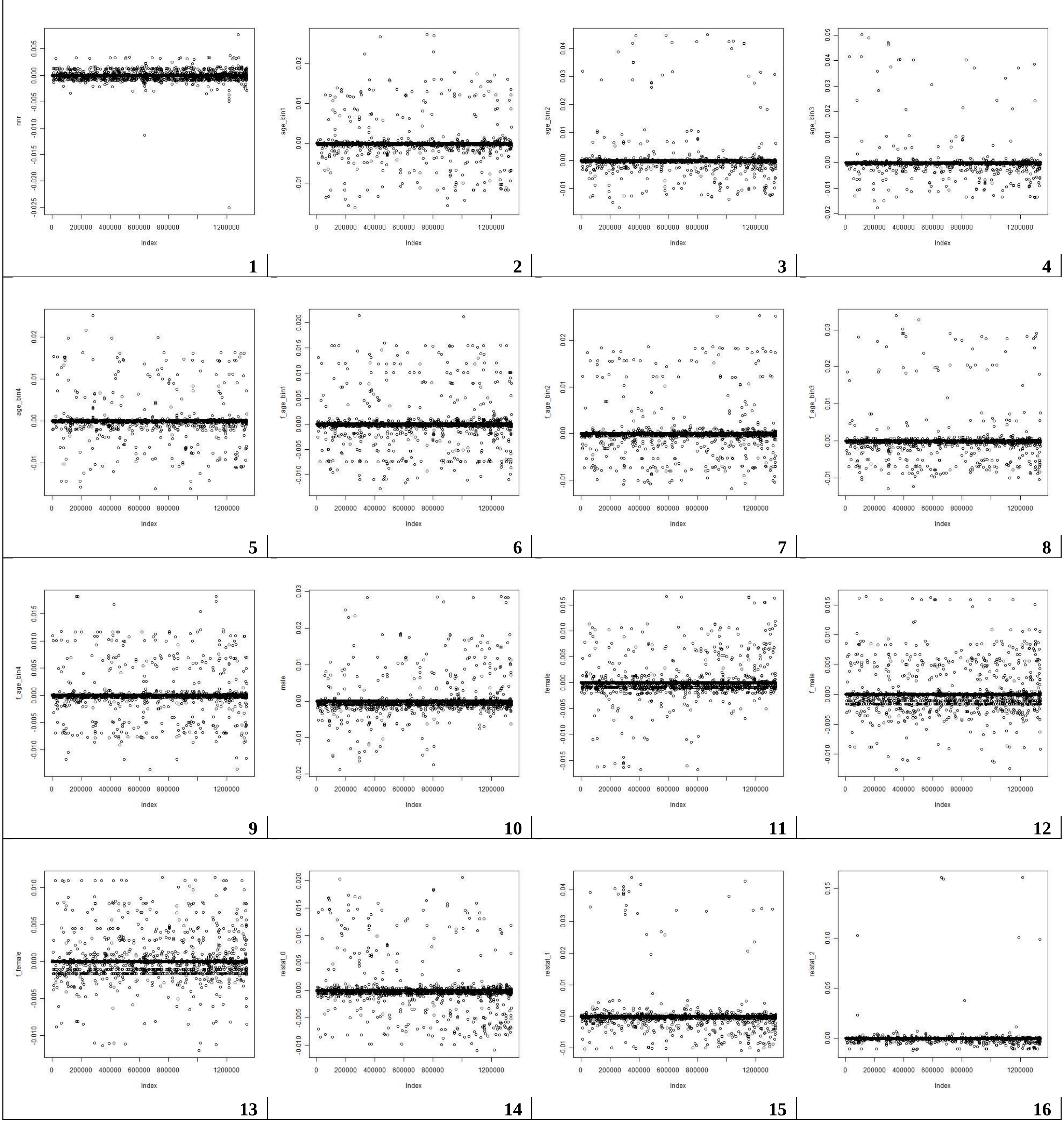


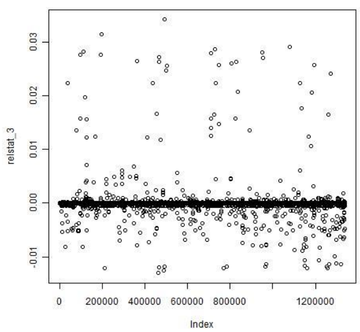
44



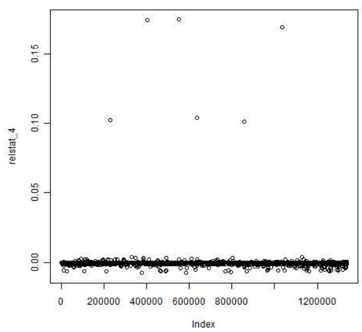
45

Fig. S10: Dfbeta Residuals for the Influence and Susceptibility Cox Proportional Hazards Model

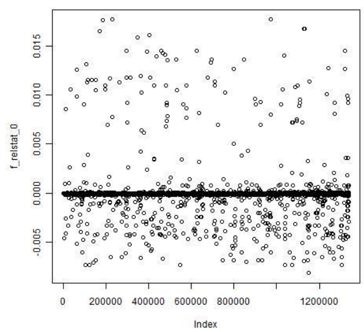




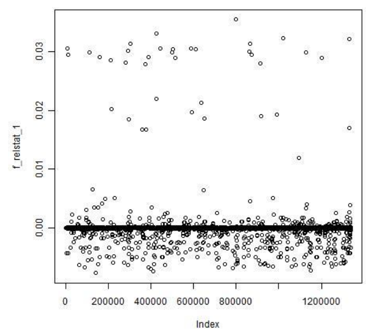
17



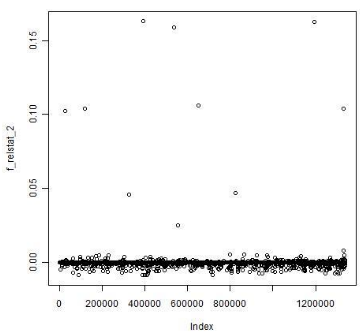
18



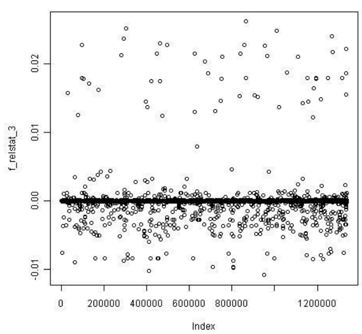
19



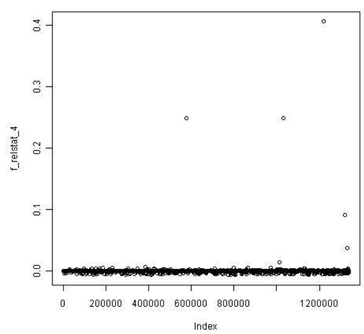
20



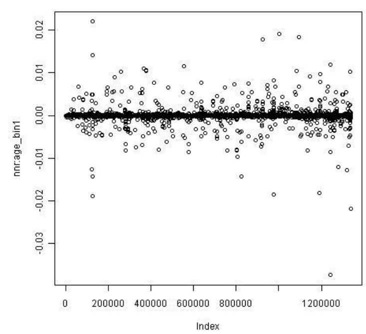
21



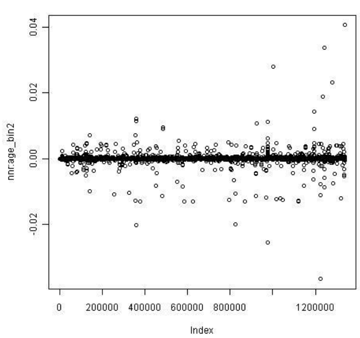
22



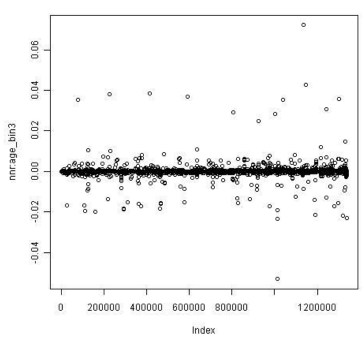
23



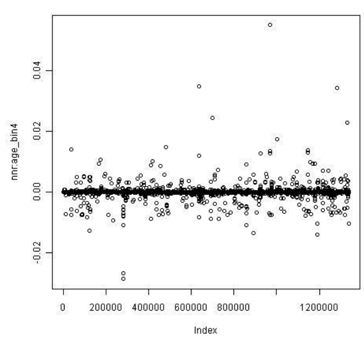
24



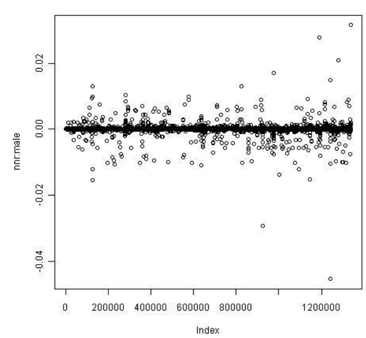
25



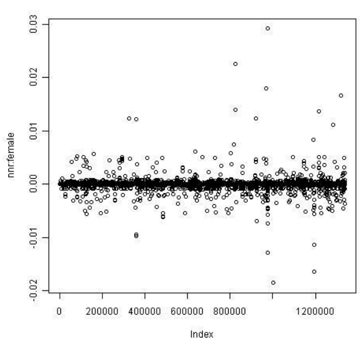
26



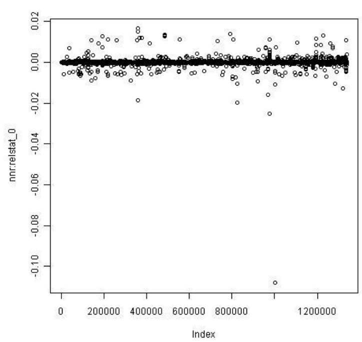
27



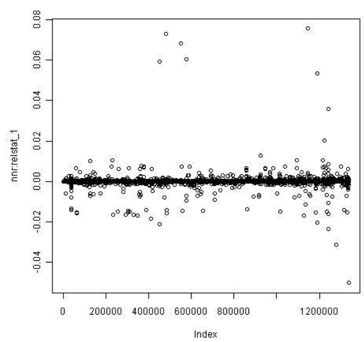
28



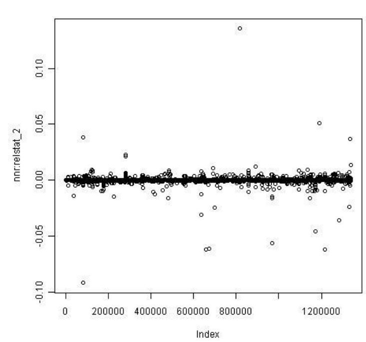
29



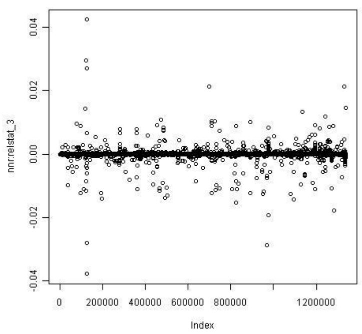
30



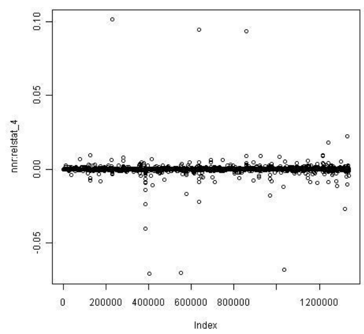
31



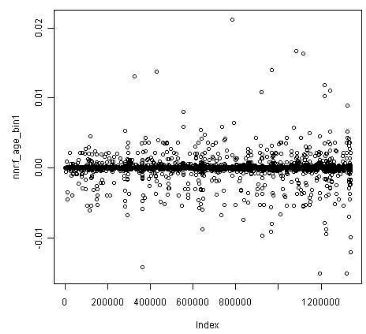
32



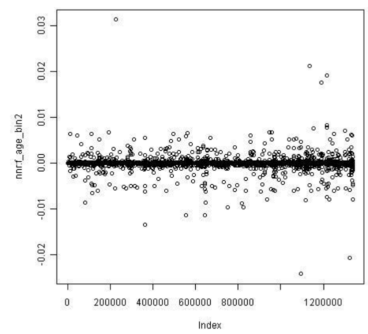
33



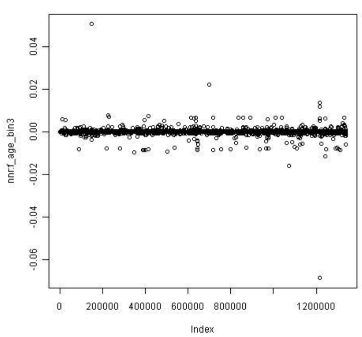
34



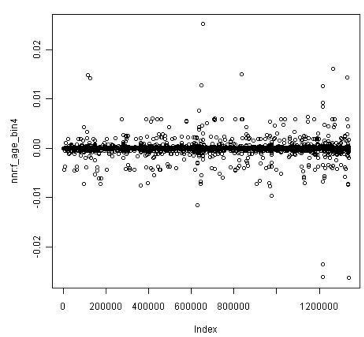
35



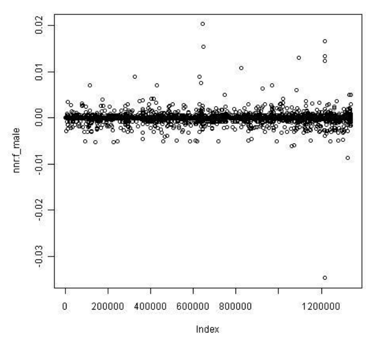
36



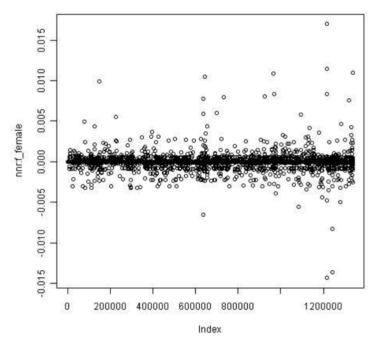
37



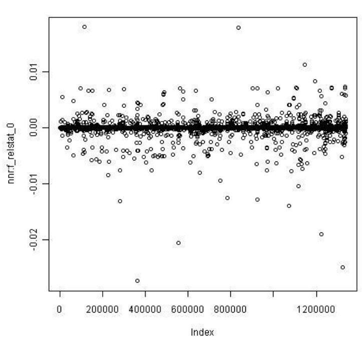
38



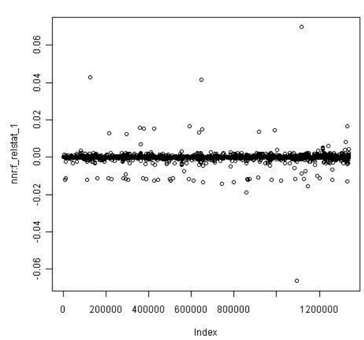
39



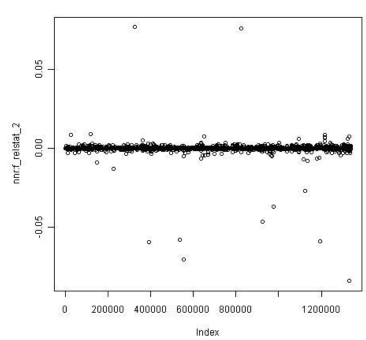
40



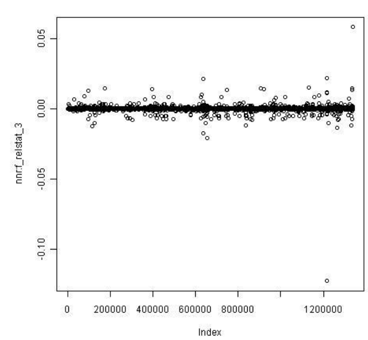
41



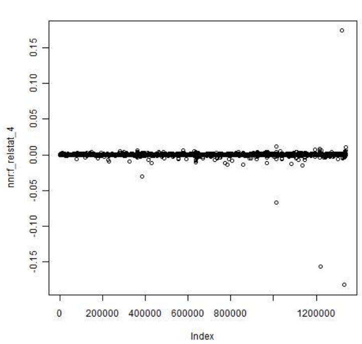
42



43

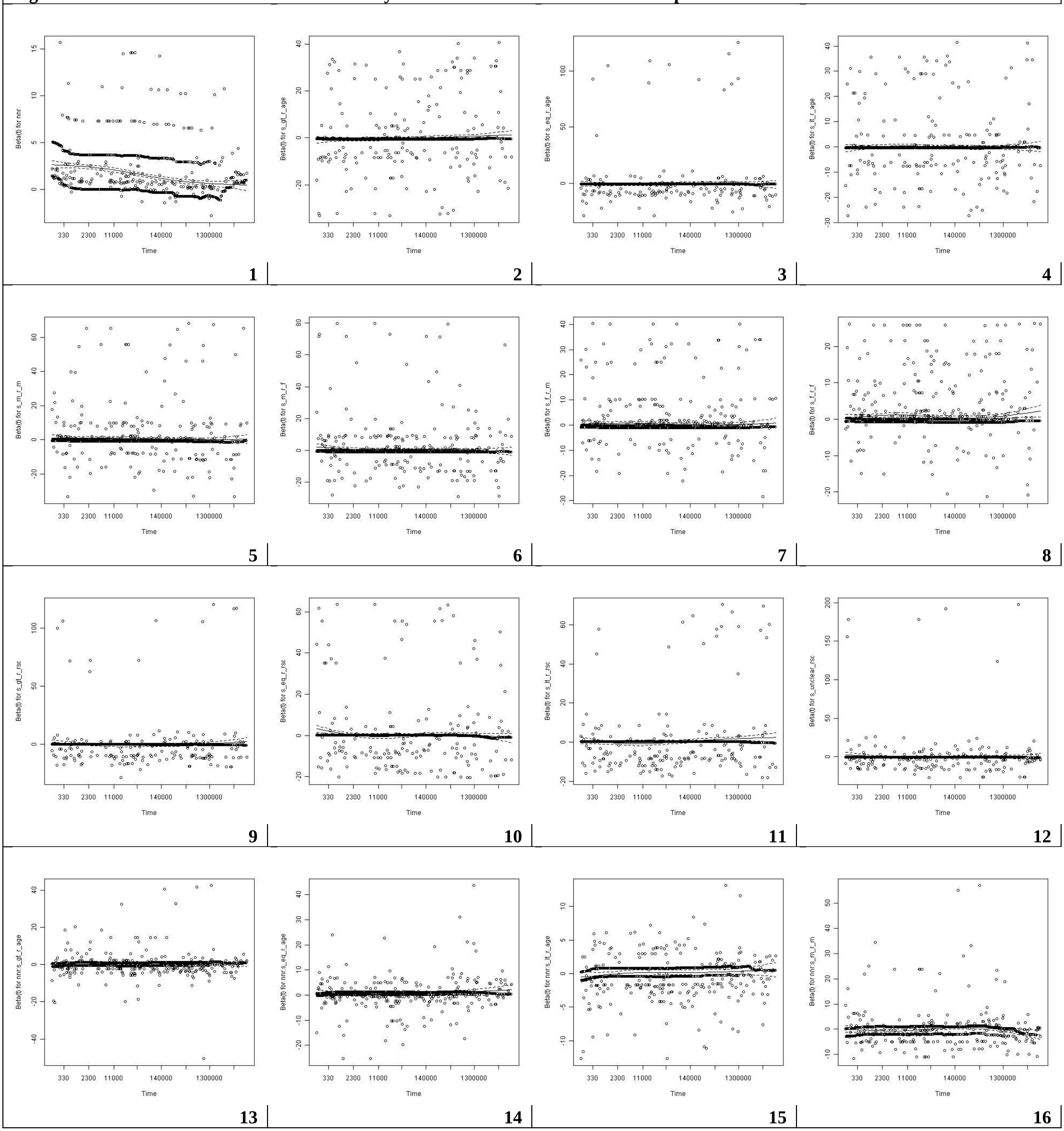


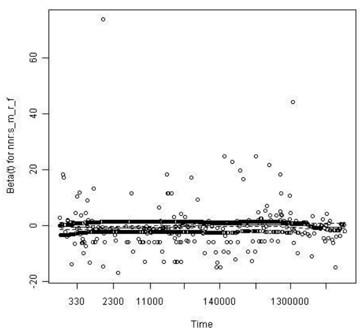
44



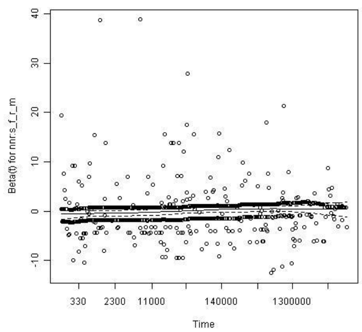
45

Fig. S11: Scaled Schoenfeld Residuals for the Dyadic Peer-to-Peer Influence Cox Proportional Hazards Model

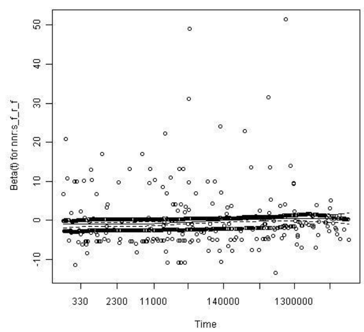




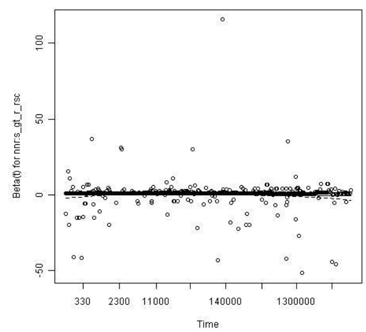
17



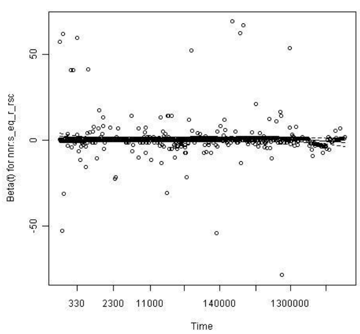
18



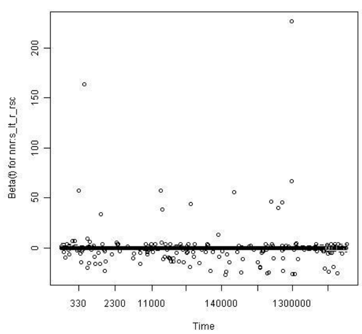
19



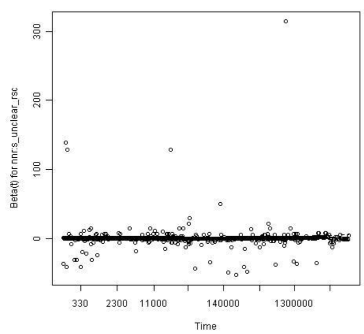
20



21

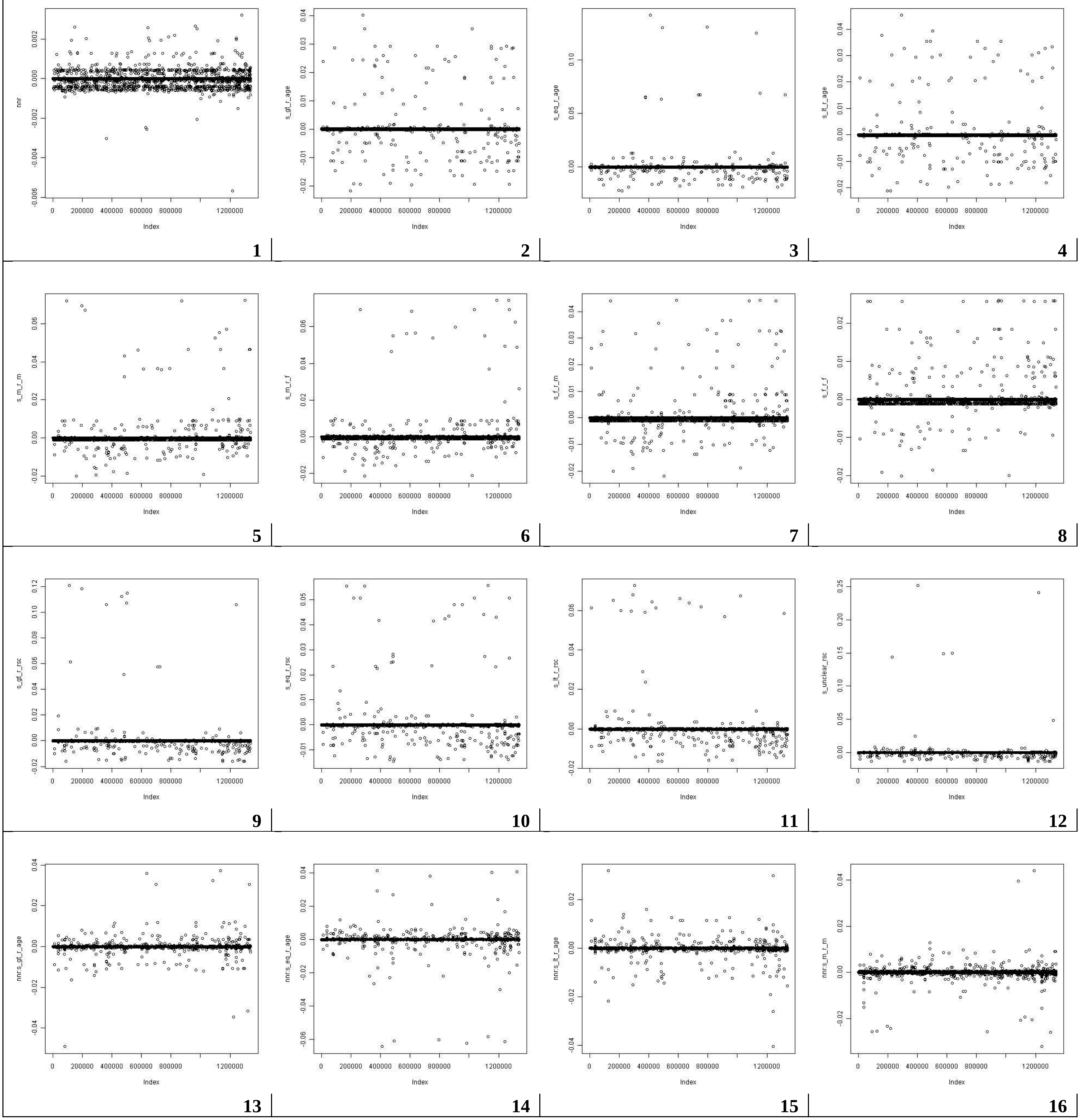


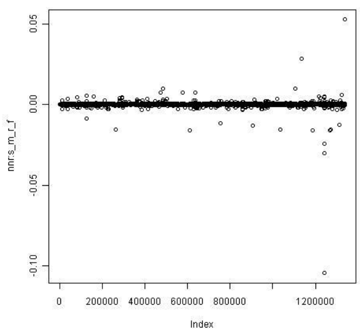
22



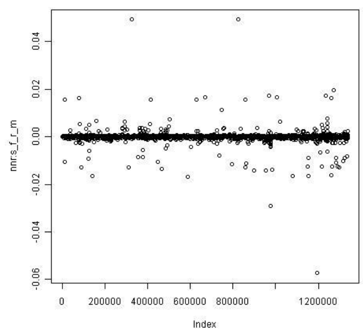
23

Fig. S12: dfbeta Residuals for the Dyadic Peer-to-Peer Influence Cox Proportional Hazards Model

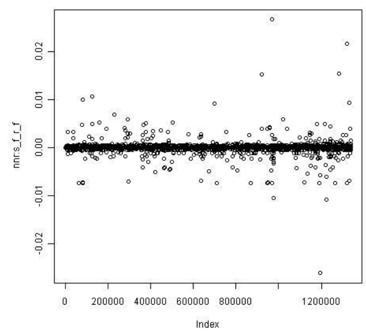




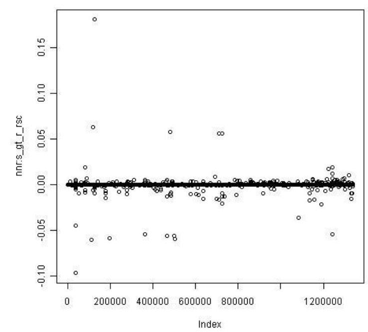
17



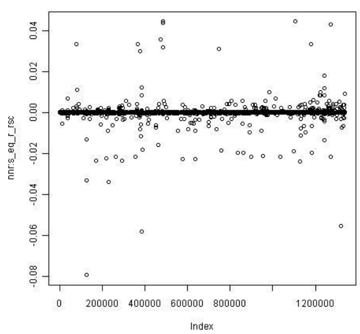
18



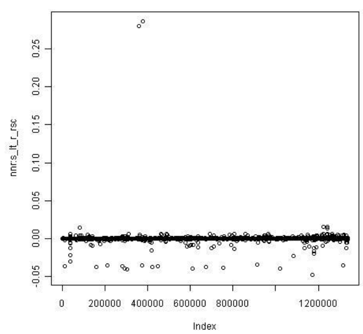
19



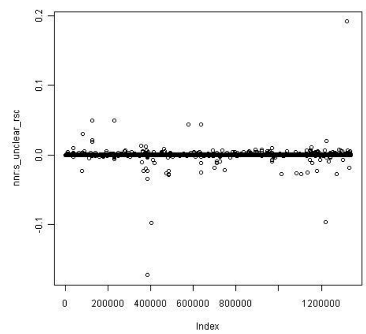
20



21



22



23