SUPPLEMENTARY ONLINE MATERIAL FOR:

THE SPREAD OF OBESITY IN A LARGE SOCIAL NETWORK OVER 32 YEARS

Nicholas A. Christakis, M.D., Ph.D., M.P.H.

James H. Fowler, Ph.D.

Supplementary Movie

A movie generated with SoNIA [1] showing the appearance and disappearance of ties among the nodes that form the largest connected subcomponent of the FHS Network is available separately. The movie documents the longitudinal change in both network topology and in attributes of the constituent individuals (*i.e.*, their BMI). Only non-genetic ties are shown in this movie (*i.e.*, friends and spouses). The movie also indicates when and to what extent the individuals (the nodes) gain and lose weight. Births and death (indicated by the appearance and disappearance of nodes) and the ties that arise or disappear as a result are shown with daily follow-up and precision; ties that arise for other reasons (*e.g.*, friendships, marriages) are noted on the date they are observed as noted on exam waves. Weights are also captured on the date of examination. Ties to immediate neighbors are not shown in this rendition. Node border indicates gender (red=female, blue=male) and arrow color denotes relation (purple=friend, gray=spouse). Node color indicates obesity (yellow=BMI>30) and node size is proportional to BMI. The date, in years and days, is shown in the upper left hand corner as time progresses.

Logistic regression models described in text

Table S1: Association of Alter Obesity and Ego Obesity, Model Set 1

	Alter Type									
	Ego-Perceived Friend	Mutual Friend	Alter-Perceived Friend	Same Sex Friend	Opposite Sex Friend	Spouse	Sibling	Same Sex Sibling	Opposite Sex Sibling	Immediate Neighbor
Alter Currently Obese	0.52	1.19	0.11	0.62	-0.29	0.37	0.40	0.53	0.28	0.32
	(0.23)	(0.33)	(0.28)	(0.24)	(0.62)	(0.15)	(0.09)	(0.13)	(0.13)	(1.50)
Alter Previously Obese	-0.62	-1.25	-0.02	-0.72	-0.01	-0.05	0.13	0.03	0.23	-0.06
	(0.25)	(0.35)	(0.29)	(0.27)	(0.55)	(0.16)	(0.09)	(0.14)	(0.13)	(0.45)
Ego Previously Obese	4.37	4.35	4.49	4.38	4.58	4.43	4.35	4.48	4.23	-0.60
	(0.18)	(0.31)	(0.22)	(0.19)	(0.47)	(0.10)	(0.12)	(0.14)	(0.14)	(0.49)
Wave 3	0.43	0.21	-0.20	0.32	1.75	0.34	0.11	0.12	0.11	4.38
	(0.28)	(0.48)	(0.36)	(0.28)	(1.24)	(0.15)	(0.17)	(0.18)	(0.21)	(0.49)
Wave 4	0.25	0.25	0.21	0.31	0.18	0.39	0.06	0.02	0.10	-0.34
	(0.23)	(0.37)	(0.35)	(0.23)	(1.24)	(0.12)	(0.14)	(0.15)	(0.17)	(0.70)
Wave 5	0.75	0.61	0.12	0.64	1.91	0.56	0.35	0.37	0.32	-0.60
	(0.25)	(0.44)	(0.33)	(0.26)	(1.20)	(0.13)	(0.15)	(0.17)	(0.17)	(0.66)
Wave 6	0.90	1.07	0.59	0.92	1.30	0.60	0.48	0.41	0.55	0.88
	(0.26)	(0.44)	(0.38)	(0.27)	(1.31)	(0.14)	(0.15)	(0.18)	(0.18)	(0.43)
Wave 7	0.79	1.03	0.00	0.71	1.87	0.47	0.25	0.20	0.30	0.77
	(0.27)	(0.48)	(0.37)	(0.28)	(1.23)	(0.14)	(0.16)	(0.18)	(0.19)	(0.60)
Ego's Age	-0.01	-0.02	-0.01	-0.01	-0.01	-0.01	0.00	0.00	0.00	1.10
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.48)
Ego Female	-0.09	-0.05	-0.18	-0.11	0.09	-0.13	0.02	-0.01	0.05	-0.01
	(0.12)	(0.22)	(0.18)	(0.13)	(0.39)	(0.07)	(0.08)	(0.10)	(0.10)	(0.02)
Ego's Years of Education	-0.01	0.02	0.05	0.01	-0.13	-0.02	-0.04	-0.04	-0.03	-0.07
	(0.03)	(0.05)	(0.04)	(0.03)	(0.08)	(0.02)	(0.02)	(0.02)	(0.02)	(0.29)
Constant	-2.40	-2.42	-2.92	-2.52	-1.94	-2.29	-2.21	-2.08	-2.35	-0.20
	(0.57)	(1.00)	(0.81)	(0.58)	(1.85)	(0.33)	(0.39)	(0.46)	(0.44)	(0.07)
Deviance	262	84	155	231	30	771	1571	768	802	63
Null Deviance	606	186	368	529	77	1803	3571	1821	1749	150
N	3504	1085	2090	3064	440	10485	20564	10300	10264	908

Coefficients and standard errors in parenthesis for logistic regression of ego obesity (1=obese, 0=not obese) on covariates shown in first column. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "friend" in the previous and current period). Models were estimated using a general estimating equation with clustering on the ego and an independent working covariance structure.[2,3] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[4]

Table S2: Association of Alter Obesity and Ego Obesity, Model Set 2

	Alter Type									
	Male Friends	Female Friends	Male Mutual Friends	Female Mutual Friends	Brothers	Sisters	Ego Brother Alter Sister	Ego Sister Alter Brother	Ego Husband Alter Wife	Ego Wife Alter Husband
Alter Currently Obese	0.84	0.30	1.28	1.09	0.44	0.61	0.31	0.26	0.44	0.34
	(0.29)	(0.44)	(0.61)	(0.33)	(0.19)	(0.17)	(0.18)	(0.18)	(0.21)	(0.22)
Alter Previously Obese	-0.90	-0.47	-1.15	-1.41	0.06	-0.01	0.25	0.20	-0.19	0.06
	(0.28)	(0.51)	(0.61)	(0.43)	(0.21)	(0.19)	(0.19)	(0.19)	(0.23)	(0.22)
Ego Previously Obese	4.23	4.61	4.37	4.41	4.26	4.74	4.31	4.15	4.26	4.67
	(0.24)	(0.30)	(0.45)	(0.46)	(0.19)	(0.20)	(0.20)	(0.20)	(0.13)	(0.16)
Wave 3	0.45	0.15	0.62	-0.26	0.35	-0.12	0.29	-0.07	0.63	-0.01
	(0.38)	(0.41)	(0.66)	(0.65)	(0.28)	(0.24)	(0.30)	(0.28)	(0.20)	(0.21)
Wave 4	0.48	0.09	0.70	-0.26	0.04	0.00	0.13	0.08	0.66	0.07
	(0.32)	(0.33)	(0.50)	(0.52)	(0.23)	(0.20)	(0.27)	(0.22)	(0.17)	(0.17)
Wave 5	0.69	0.56	1.03	0.14	0.34	0.40	0.22	0.42	0.84	0.25
	(0.35)	(0.39)	(0.50)	(0.65)	(0.24)	(0.23)	(0.23)	(0.25)	(0.18)	(0.20)
Wave 6	0.80	0.97	1.34	0.72	0.53	0.29	0.67	0.45	0.90	0.25
	(0.36)	(0.39)	(0.62)	(0.57)	(0.24)	(0.26)	(0.26)	(0.25)	(0.18)	(0.22)
Wave 7	1.16	0.21	1.56	0.46	0.35	0.04	0.28	0.33	0.78	0.12
	(0.39)	(0.40)	(0.65)	(0.67)	(0.25)	(0.27)	(0.28)	(0.27)	(0.20)	(0.21)
Ego's Age	-0.02	0.00	-0.02	-0.01	-0.01	0.00	0.00	0.00	-0.01	0.00
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ego's Years of Education	0.02	-0.02	0.01	0.02	-0.05	-0.03	-0.03	-0.04	-0.02	-0.02
	(0.04)	(0.05)	(0.06)	(0.10)	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)
Constant	-2.44	-2.60	-2.48	-2.30	-1.82	-2.32	-2.39	-2.25	-2.19	-2.45
	(0.80)	(0.82)	(1.38)	(1.50)	(0.64)	(0.61)	(0.60)	(0.65)	(0.43)	(0.52)
Deviance	118	112	40	44	385	382	387	414	422	348
Null Deviance	269	257	98	86	872	947	893	856	968	827
N	1431	1633	521	552	4736	5564	5093	5171	5199	5286

Coefficients and standard errors in parenthesis for logistic regression of ego obesity (1=obese, 0=not obese) on covariates shown in first column. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "friend" in the previous and current period and both are males). Models were estimated using a general estimating equation with clustering on the ego and independent working covariance structure.[2,3] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[4]

Additional Statistical Information and Sensitivity Analyses

The models in Tables S1 and S2 provide parameter estimates in the form of beta coefficients, whereas the results reported in the text and in Figures 3 and 4 of the paper are in the form of odds ratios, which are related to the exponentiated coefficients. The key coefficients are the effect of alter obesity at t+1. In many of the models in Tables S1 and S2 related specifically to friendship ties, the coefficient for alter obesity at t is negative. Given the fact that the models also control for alter obesity at $t+1$ and for ego obesity at t and $t+1$, this may be interpreted as a tendency for *heterophily*, or the tendency of egos to nominate alters who are not of the same obesity status as egos (the "Laurel and Hardy" effect); models of familial ties tend not to have any negative coefficients, as would be expected. As shown in Table S3, there is no evidence for heterophily among friends when it comes to smoking behavior.

The other regression coefficients have mostly the expected effects, such that, for example, less educated individuals are more likely to be obese. As indicated, the models in the foregoing tables include wave fixed effects, which, combined with age at baseline, account for the aging of the population over the 32 years.

We estimated these models on the ego/alter pair types described. We also estimated models that treated the pair type as a factor variable that was interacted with the BMI variables; these models did not yield substantively different results.

The sample size, *N*, shown in Tables S1 and S2 reflects the total number of all such ties, with multiple observations for each tie if it was observed in more than one wave, and allowing for the possibility that a given person can have multiple ties. Hence, for example, there are 20,564 observations of ego-alter sibling ties across all seven waves in the network.

We explored the sensitivity of our results to model specification by conducting numerous other analyses (not shown here) each of which had various strengths and limitations, but none of which yielded substantially different results than those presented here. We specified models in which we lagged the alter's weight status by more than one period. We modeled how changes in the alter's weight status between two periods affected ego's weight status in the subsequent period. Although we identified only a single friend for most of the egos, we studied how multiple observations on some egos affected the standard errors of our models. Huber-White sandwich estimates with clustering on the egos yielded very similar standard errors. And we specified models that included a fixed effect for each ego (which drops all observations of egos with a single friend since they have no variation), thus controlling for all time-invariant attributes of the egos, such as their genes.

The Kamada-Kawai algorithm used to prepare the images in Figures 1 and 2 in the paper generates a matrix of shortest network path distances from each node to all other nodes in the network and repositions nodes so as to reduce the sum of the difference between the plotted distances and the network distances.[5]

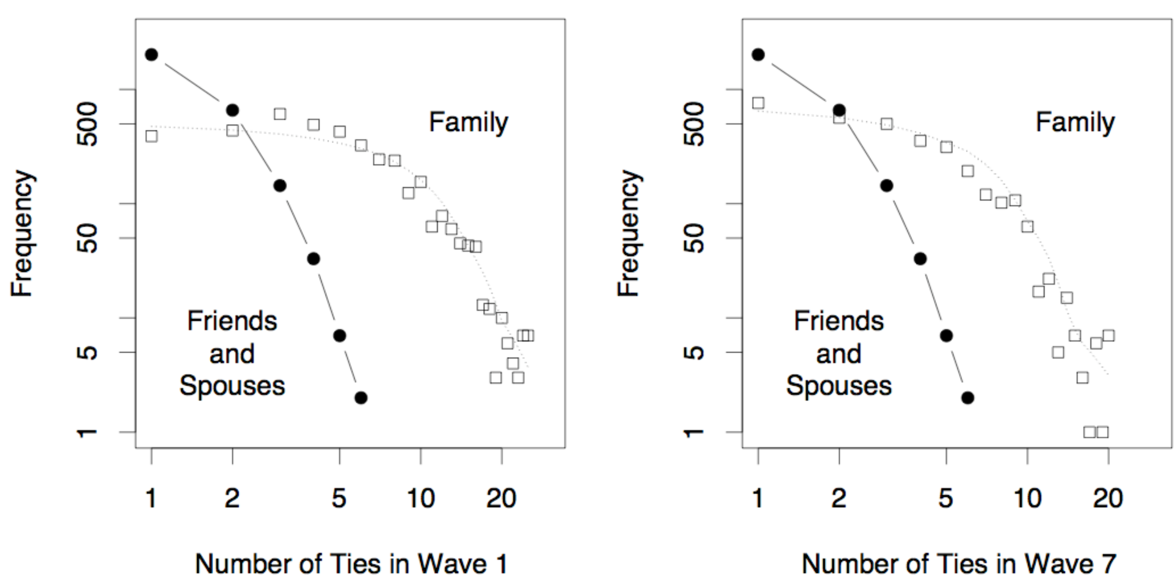
Degree Distribution in the Framingham Heart Study Social Network

We evaluated whether our data conform to theoretical network models that have been developed for metabolic, communication, and other networks, such as the small-world,[6] scale-free,[7] and hierarchical types.[8]

The *degree* of a node indicates how many ties it has to other nodes. A common procedure used in many network studies is to generate the *degree distribution*, a logarithmic histogram showing the frequency of nodes that are sparsely or densely connected to the network. Many social networks exhibit the property that most nodes have low degree while a few have very high degree, with exponential decay in the distribution. In particular, the distributions for some networks decay at a constant exponential rate, indicating they are "scale free" and may have resulted from a preferential attachment process in which new nodes are more likely to attach to nodes that are already attached to many nodes.[7] A straight line on the logarithmic histogram would indicate a power law in the distribution, which is characteristic of scale-free networks.

Here we present separate distributions for the friend and spouse network and family network for waves 1 and 7 (the other waves are similar). These figures show that most individuals have one or two friends and 10 or fewer family members who also participate in the Framingham Study. However, a small number of nodes are very well-connected, as observed in other social networks studied elsewhere.[7] The breadth of the degree distribution suggests that the small world network model is inappropriate because it typically produces a peaked distribution with most nodes having the same number of connections close to the average.[6] Furthermore, notice that none of these distributions conform to a power law, suggesting the network we explore may not be scale free.[8]

Figure S1: Degree Distribution of the FHS-Net



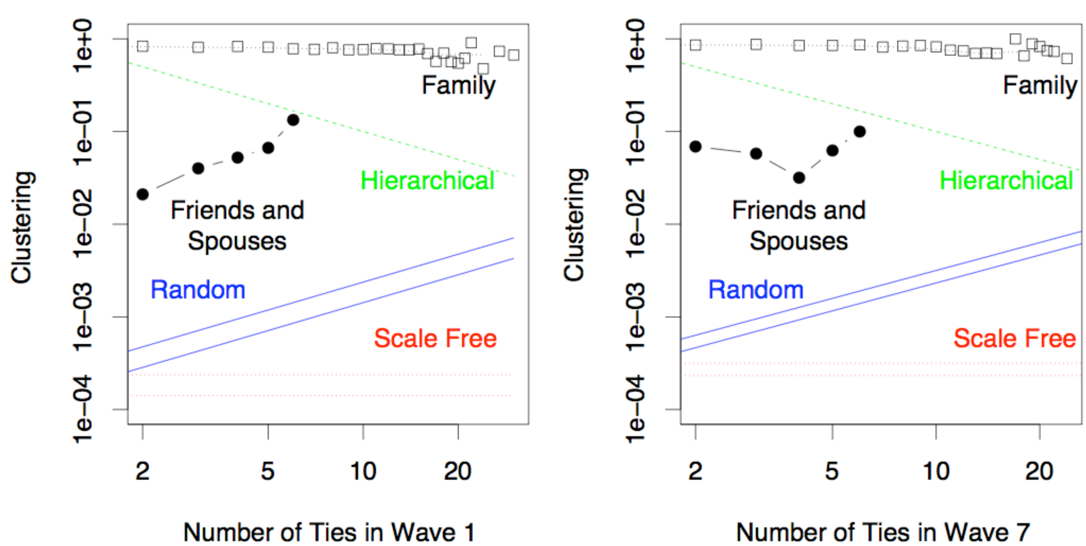
Assortativity in the Framingham Heart Study Social Network

Another property of human social networks is *assortativity*, or the tendency for well-connected individuals to be connected to one another.[9] Networks of scientific coauthorship, film actor collaboration, and company directors all exhibit a high degree of assortativity.[10] An easy way to measure assortativity is a simple Pearson correlation between ego degree and alter degree. When we do this for the Framingham network, we find very high assortativity in the degree of family relations ($r = 0.45$). This is not surprising, since many family relations are transitive by definition (*e.g.*, all siblings will have exactly the same number of siblings). However, we also find high assortativity in the number of friend and spouse relationships ($r = 0.28$). In fact, this is almost exactly the same level of assortativity found in the network of company directors. What these measures show is that the network is obviously not random -- otherwise, r would equal zero. However, the preferential attachment model also yields zero assortativity, suggesting that it cannot account for the generic process by which these friendships and family relations in the Framingham social network tend to arise.

Clustering in the Framingham Heart Study Social Network

Another property of networks is *clustering*, or the degree to which nodes that link to one another also link to the same other nodes. The *clustering coefficient* is the principal measure for this phenomenon, indicating the fraction of ego's relations that are related to one another. For example, if an ego has three friends who are all friends with one another, then that ego's clustering coefficient is 1. If however, one of these friendship ties is broken, so that A and B are friends and B and C are friends, but A and C are not, then the clustering coefficient is 2/3. The average clustering coefficient for the Framingham social network is approximately 0.66, which is consistent with many other observed social networks.

Figure S2: Clustering Coefficients in the FHS-Net



Here we show the observed average clustering coefficient in the Framingham social network as a function of degree for the friend and spouse (solid circles) and family (open squares) networks. The lines in color represent theoretical expectations based on the size of the network and the degree for the friend and spouse network (top line of each color pair) and the family network (bottom line of each color pair). These figures show that the amount of clustering observed in the family network and the friend and spouse networks is at least an order of magnitude greater than the clustering predicted by the random or scale-free network models. Although the hierarchical network model can achieve a high level of clustering, notice that the observed relationship between clustering and degree does not decay as predicted by hierarchical network models.[8] Finally, note that the small world network [6] can achieve these high levels of clustering, but as noted previously its degree distribution is typically unrealistically narrow.

Effect of Ego Connectedness

A number of studies have suggested the importance of well-connected nodes in networks for spreading processes.[9,11] We thus explored the effect of ego's degree on obesity. If well-connected individuals tend to be obese (or not), it might affect our results since these individuals by definition affect the dyadic observations of a large number of individuals. As a first cut, we pooled data across waves and conducted a Pearson's moment correlation test on obesity and degree. The results suggest there is no significant relationship between obesity and the number of friendship and spouse ties (r=0.006, 95% confidence interval -0.007, 0.019, t = 0.906, df = 21,608, p-value = 0.36). We also tried adding the number of friendship and spouse ties for both ego and alter to the statistical models, both alone and as an interaction term with alter's obesity in the current period. None of these coefficients in any of these models were significant (all p>0.23). We include these covariates in the full model of ego/friend ties below for illustration.

Effect of Smoking Behavior and of Geographic Distance between Ego and Alter

In addition to controlling for ego and alter node degree, we were interested in exploring the role of physical distance and smoking as possible factors in the influence of alter on ego. As suggested in related results in the text, physical distance between ego and alter does not appear to influence our results. When we tried adding distance and square root of distance between ego and alter, both alone and as an interaction term with alter's obesity in the current period, none of the models we tried yielded significant coefficients for the additional terms. We include the distance measure in the full model below for illustration.

We also added smoking to the model, both as a dichotomous variable indicating whether ego and alter have smoked in the last year and as a count variable of the number of cigarettes smoked per day. We included measures at both the current and previous wave. Since the coefficient for alter obesity is virtually identical to the one reported in the first table above for the "Friend" model, these results suggest that smoking is not a mediator of the interpersonal spread of obesity.

Finally, we also specified models in which each of the foregoing variables (degree, smoking, and distance) was added singly to the core model, and this did not yield different results.

Table S3: Models With Extra Controls For Smoking, Distance, and Degree

	Friends				*Siblings*			
	Coef.	**S.E.**	**Wald**	**p(>W)**	**Coef.**	**S.E.**	**Wald**	**p(>W)**
Alter Currently Obese	0.639	0.228	7.886	0.005	0.433	0.099	19.190	0.000
Alter Previously Obese	-0.775	0.242	10.210	0.001	0.092	0.109	0.711	0.399
Ego Currently Obese	4.673	0.200	546.400	0.000	4.365	0.134	1067.000	0.000
Alter Currently Smokes	-0.086	0.299	0.082	0.774	0.033	0.106	0.099	0.753
Alter Smoked in Previous Wave	0.073	0.290	0.064	0.800	-0.523	0.181	8.340	0.004
Ego Currently Smokes	-0.422	0.306	1.893	0.169	-0.057	0.107	0.281	0.596
Ego Smoked in Previous Wave	0.395	0.301	1.722	0.189	0.409	0.171	5.729	0.017
Wave 3	0.428	0.331	1.673	0.196	0.139	0.189	0.539	0.463
Wave 4	-0.017	0.290	0.003	0.953	-0.064	0.154	0.171	0.680
Wave 5	0.545	0.313	3.029	0.082	0.235	0.169	1.930	0.165
Wave 6	0.884	0.336	6.937	0.008	0.431	0.180	5.698	0.017
Wave 7	0.706	0.353	4.011	0.045	0.182	0.188	0.935	0.334
Ego's Age	-0.009	0.011	0.663	0.415	-0.005	0.006	0.812	0.368
Alter's Age	-0.007	0.009	0.612	0.434	0.002	0.006	0.108	0.742
Ego's Gender	0.018	0.224	0.007	0.935	-0.011	0.090	0.015	0.901
Alter's Gender	-0.091	0.210	0.188	0.665	-0.052	0.054	0.908	0.341
Ego's Education	0.002	0.033	0.004	0.947	-0.039	0.023	2.892	0.089
Alter's Education	-0.039	0.035	1.224	0.269	-0.021	0.017	1.581	0.209
Ego's Family Ties	-0.015	0.019	0.588	0.443	-0.036	0.030	1.427	0.232
Alter's Family Ties	0.000	0.021	0.000	0.997	0.039	0.041	0.916	0.339
Ego's Inward Friendship Ties	0.047	0.085	0.300	0.584	0.109	0.060	3.228	0.072
Alter's Inward Friendship Ties	0.044	0.084	0.272	0.602	-0.015	0.044	0.108	0.743
Ego's Outward Friendship Ties	-0.113	0.138	0.679	0.410	-0.047	0.072	0.422	0.516
Alter's Outward Friendship Ties	-0.100	0.108	0.854	0.356	-0.014	0.057	0.060	0.807
Geographic Distance Between Ego and Alter (miles)	0.000	0.000	0.144	0.704	0.000	0.000	3.232	0.072
Constant	-1.481	0.882	2.815	0.093	-1.752	0.522	11.290	0.001
Deviance	194				1266			
Null Deviance	448				2897			
N	2747				16535			

Logistic regression of ego obesity (1=obese, 0=not obese) on covariates shown in first column. Coefficients, standard errors, and results of a Wald test for significance are shown. Observations for this model are restricted to friends named by egos. Models were estimated using a general estimating equation with clustering on the ego and independent covariance structure.[2,2] Models with an exchangeable correlation structure yielded poorer fit. Models with the natural logarithm of miles did not yield substantively different results. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[4]

References

1. Moody, J, McFarland DA, and Bender-deMoll S. Visualizing Network Dynamics. *American Journal of Sociology* 2005; 110: 1206-41.
2. Liang, K.Y. and Zeger, S.L. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; 73: 13-22.
3. Schildcrout JS. Regression analysis of longitudinal binary data with time-dependent environmental covariates: bias and efficiency. *Biostatistics* 2005 6(4):633-652.
4. Wei, P. Goodness-of-fit Tests for GEE with Correlated Binary Data. *Scandinavian Journal of Statistics* 2002; 29 (1), 101110
5. Kamada T and Kawai S. An algorithm for drawing general undirected graphs. *Information Processing Letters* 1989; 31: 7-15.
6. Watts, DJ and Strogatz SH. Collective Dynamics of Small World Networks. *Nature* 1998; 393(6684): 490-410.
7. Barabasi AL and Albert R. Emergence of scaling in random networks. *Science* 1999; 286: 509-512.
8. Ravasz E and Barabasi AL. Hierarchical organization in complex networks. *Physical Review E* 2003; 67: 026112.
9. Pastor-Satorras et al. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett*. 2001; 87: 258701.
10. Newman MEJ. Assortative Mixing in Networks. *Phys. Rev. Lett.* 2002; 89, 208701.
11. Albert, R. Jeong, H. Barabasi, AL. Error and attack tolerance of complex networks. *Nature* 2000;406(6794):378-82.