Using Social Network Analysis to Examine the Collective Dynamics of Smoking Behavior

Senior Statistics Colloquium

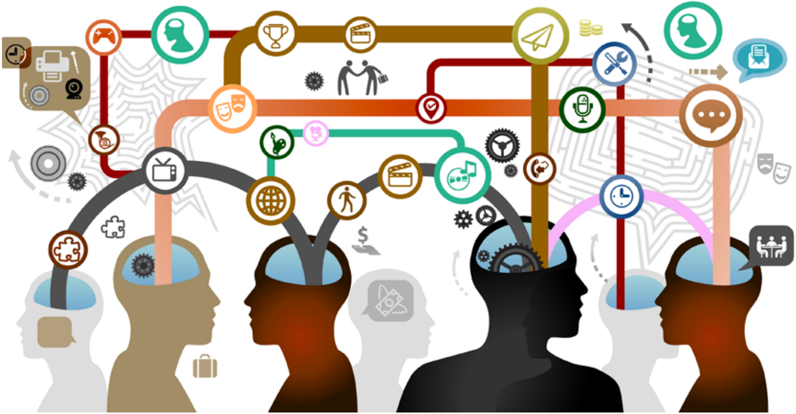
Isabel Arvelo

February 15, 2023

Table of Contents

Table of Contents

Motivation



Motivation

- The prevalence of smoking declined from 45% to 21% between 1965 and 2008
- Examining the collective dynamics of smoking behavior within a social network can inform clinical and public health interventions to reduce and prevent smoking

Table of Contents

Definition: A network is a collection of elements and their inter-relations.

Definition: A graph is an ordered pair G=(V,E) consisting of a nonempty set V (called the vertices) and a set E (called the edges) of two-element subsets of V.

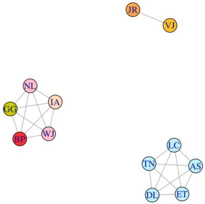


Figure: Visualization of Divisions of Class of 2023 Statistics Majors' Double Majors

Social Network Analysis Terminology

Definition: The graph theoretic distance d(u,v) between two vertices u and v of a finite graph is the minimum length of the paths connecting them (i.e., the length of a graph geodesic)

Definition: A graph is a connected graph if, for each pair of vertices, there exists at least one single path which joins them.

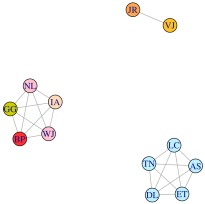


Figure: Visualization of Divisions of Class of 2023 Statistics Majors' Double Majors

Social Network Analysis Terminology

Definition: Network topology refers to the manner in which the links and nodes of a network are arranged to relate to each other.

Definition: An ego is an individual "focal" node

Definition: An alter is any persons to whom egos are linked

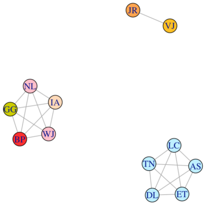


Figure: Visualization of Divisions of Class of 2023 Statistics Majors' Double Majors

Social Network Analysis Terminology

Definition: Network cohesion is the extent to which subsets of vertices are cohesive, or 'stuck together', with respect to the relation defining edges in the network graph

Definition: A cluster refers to a tightly connected community in which adjacent nodes of a node are connected and share a common behavior

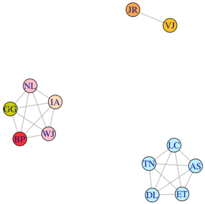


Figure: Visualization of Divisions of Class of 2023 Statistics Majors' Double Majors

Clustering

How can we rigorously assess if there is clustering happening?

- Compare the whole observed network to simulated "random networks"
  - Definition: A random network has the same network topology and the overall prevalence of the response variable of interest as the observed network, but with the incidence of the response randomly distributed across the nodes
- If clustering is occurring, then the probability that a contact exhibits a certain behavior, given that a subject does, should be higher in the observed network than in the random networks

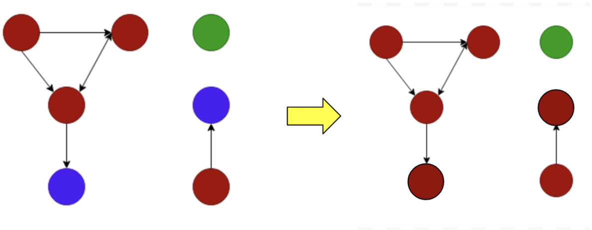
Explanations for Clustering

Main explanations for clustering

1. Induction/Assimilation
2. Homophily
3. Confounding

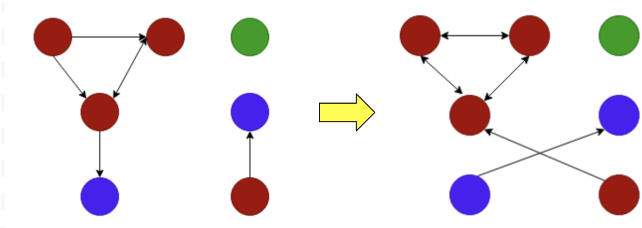
Explanations for Clustering

Induction(Assimilation): Subjects influence each other, so the response can be thought of as contagious



Explanations for Clustering

Homophily: Subjects choose to associate with contacts with similar behaviors



Explanations for Clustering

Confounding: Subjects share a common unobserved contemporaneous exposure that causes their behaviors to covary

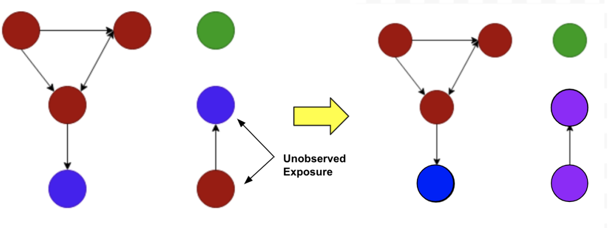
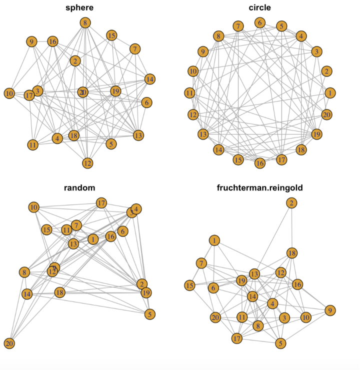


Table of Contents

Pleasing Graph Layouts

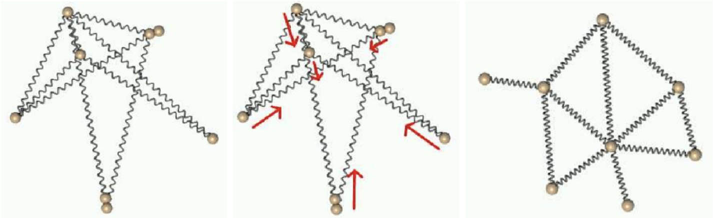
What is the best way to effectively communicate the desired relational information in G?

- Infinitely many ways to arrange a candidate set of points and curves on paper to represent a graph G
- We can optimize aesthetically pleasing graph layouts using only information contained within the structure of the graph itself



Force-directed Graph Drawing

Definition: Force-directed graph drawing algorithms are a class of algorithms for drawing simple, undirected graphs in an aesthetically pleasing way using expressions motivated by those found in Physics.



Vertex placement is chosen which minimizes the total system energy, E

Kamada and Kawai: Spring Model

Let p_1, p_2, ... p_n be the particles in a plane mutually connected by springs corresponding to the vertices v_1, v_2, ..., $v_n \in V$
Let l_{ij} correspond to the desirable length of the spring between p_i and p_j in the drawing.
Let k_{ij} be the strength of the spring between p_i and p_j

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2} k_{ij}(|p_i - p_j| - l_{ij})^2 \tag{1}$$

Pleasing layouts can be obtained by decreasing E, with the best layout minimizing E.

Kamada and Kawai: Spring Model

Let p_1, p_2, ... p_n be the particles in a plane mutually connected by springs corresponding to the vertices v_1, v_2, ..., $v_n \in V$

Let l_{ij} correspond to the desirable length of the spring between p_i and p_j in the drawing.

Let k_{ij} be the strength of the spring between p_i and p_j

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2} k_{ij}(|p_i - p_j| - l_{ij})^2 \tag{1}$$

Pleasing layouts can be obtained by decreasing E, with the best layout minimizing E.

Kamada and Kawai: Spring Model

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2} k_{ij} (|p_i - p_j| - l_{ij})^2 \tag{1}$$

Let L be the desirable length of a single edge in the display plane
Let d_{ij} be the theoretic or geodesic distance between two vertices p_i and p_j

$$l_{ij} = L x d_{ij} \tag{2}$$

Let L_0 be the length of a side of the display square area
Let $max\ d_{ij}$ be the diameters of the graph (distance between the farthest pair of vertices)

$$L = L_0 / max d_{ij}, i < j \tag{3}$$

Kamada and Kawai: Spring Model

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2} k_{ij} (|p_i - p_j| - l_{ij})^2 \tag{1}$$

K is a constant and the parameters l_{ij} and k_{ij} are symmetric

$$k_{ij} = K / d_{ij}^2 \tag{4}$$

Kamada and Kawai: Local minimization of global energy

The position of a particle in a plane is expressed by X and Y coordinate values. Let (x_1, y_1), (x_2, y_2), ... ,(x_n, y_n) be the coordinate variables of particles p_1, p_2, ... p_n respectively. Then, the energy E defined as (1) is rewritten by using these 2n variables

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2} k_{ij} \{(x_i - x_j)^2 + (y_i - y_j)^2 + l_{ij}^2 - 2 l_{ij} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\} \tag{5}$$

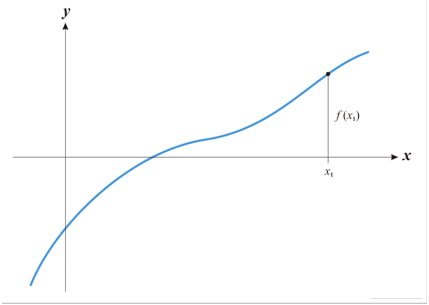
Kamada and Kawai: Local minimization of global energy

It is difficult to compute the minimum of $E(x_1, x_2, \ldots, y_1, y_2, \ldots)$, so we instead approximate a local minimum using the Newton-Raphson method.

<u>Definition</u>: Newton-Raphson method is a root-finding algorithm that produces successively better approximations to the roots of a real-valued function. It uses the fact that a continuous and differentiable function can be approximated by a straight line tangent to it.

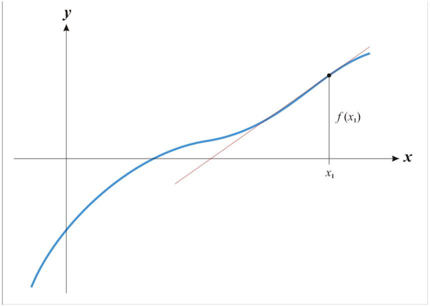
Kamada and Kawai: Local minimization of global energy

Newton-Raphson Method



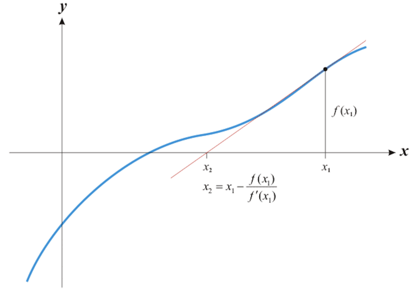
Kamada and Kawai: Local minimization of global energy

Newton-Raphson Method



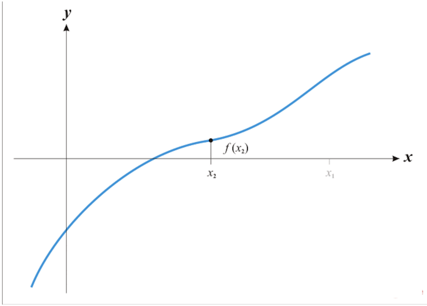
Kamada and Kawai: Local minimization of global energy

Newton-Raphson Method



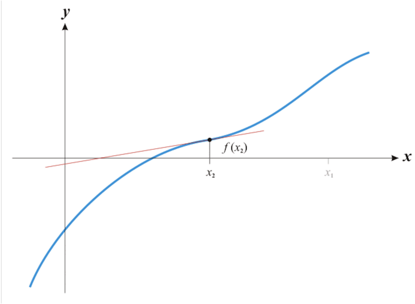
Kamada and Kawai: Local minimization of global energy

Newton-Raphson Method

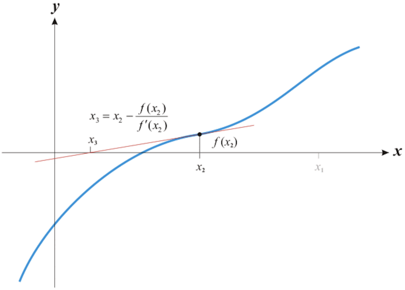


Kamada and Kawai: Local minimization of global energy

Newton-Raphson Method

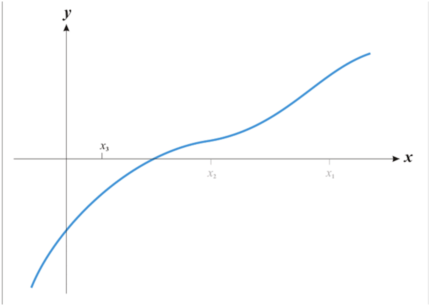


Newton-Raphson Method



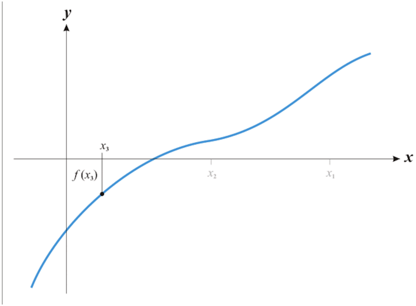
Kamada and Kawai: Local minimization of global energy

Newton-Raphson Method



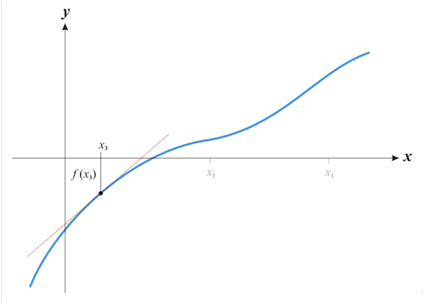
Kamada and Kawai: Local minimization of global energy

Newton-Raphson Method



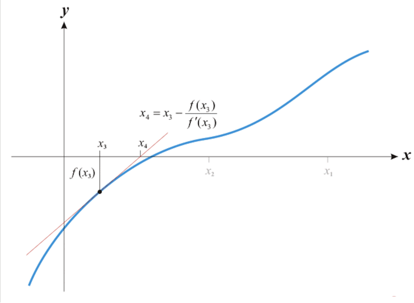
Kamada and Kawai: Local minimization of global energy

Newton-Raphson Method



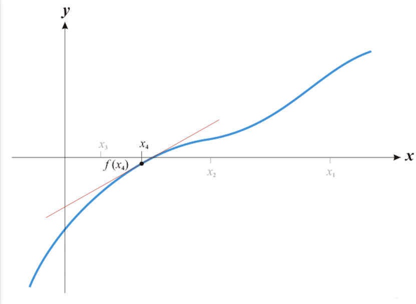
Kamada and Kawai: Local minimization of global energy

Newton-Raphson Method



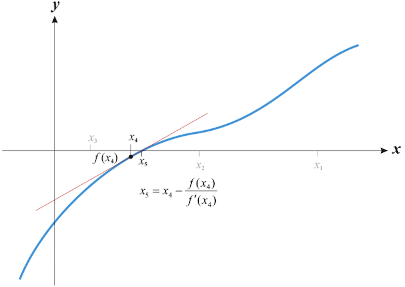
Kamada and Kawai: Local minimization of global energy

Newton-Raphson Method



Kamada and Kawai: Local minimization of global energy

Newton-Raphson Method



Kamada and Kawai: Local minimization of global energy

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2} k_{ij} \{(x_i - x_j)^2 + (y_i - y_j)^2 + l_{ij}^2 \\ -2 l_{ij} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\} \tag{6}$$

Equation 7 corresponds to the dynamic state in which the forces of all springs are balanced

$$\frac{\partial E}{\partial x_m} = \frac{\partial E}{\partial y_m} = 0 for 1 < m < n \tag{7}$$

Kamada and Kawai: Local minimization of global energy

$$\frac{\partial E}{\partial x_m} = \sum_{i \neq m} k_{mi} \left[(x_m - x_i) - \frac{l_{mi}(x_m - x_i)}{\{(x_m - x_i)^2 + (y_m - y_i)^2)\}^{\frac{1}{2}}} \right] \tag{8}$$

$$\frac{\partial E}{\partial y_m} = \sum_{i \neq m} k_{mi} \left[(y_m - y_i) - \frac{l_{mi}(x_m - x_i)}{\{(x_m - x_i)^2 + (y_m - y_i)^2)\}^{\frac{1}{2}}} \right] \tag{9}$$

Kamada and Kawai: Local minimization of global energy

We must solve these 2n simultaneous non-linear equations of 7, but they cannot be directly solved by using a 2n-dimensional Newton-Raphson method, because they are not independent of one another.

$$\frac{\partial E}{\partial x_m} = \frac{\partial E}{\partial y_m} = 0 \text{ } for \text{ } 1 < m < n \tag{7}$$

What do we do now?

- Move one particle $P_m(x_m, y_m)$ to its stable point at a time, freezing the other particles.
- Obtain a local minimum that satisfies 7 by using a two-dimensional Newton-Raphson method to minimize E as a function of only x_m and y_m and iterate this step.

Kamada and Kawai: Local minimization of global energy

In each step, we choose the particle that has the largest value of Δ_m which is defined as

$$\Delta_m = \sqrt{\{\frac{\partial E}{\partial x_m}\}^2 + \{\frac{\partial E}{\partial y_m}\}^2} \tag{10}$$

Starting from the current position $(x_m^{(0)}, y_m^{(0)})$, the following step is iterated for t$=$ 0, 1, 2 ...

$$\begin{aligned} x_m^{(t+1)} &= x_m^{(t)} + \delta x \\ y_m^{(t+1)} &= y_m^{(t)} + \delta y \end{aligned} \tag{11}$$

The iteration 11 terminates when the value of Δ_m at $(x_m^{(t)}, y_m^{(t)})$ becomes small enough.

Kamada and Kawai: Algorithm

First steps of graph drawing algorithm:

- compute d_{ij} for $1 \leq i \neq j \leq n$
- compute l_{ij} for $1 \leq i \neq j \leq n$
- compute k_{ij} for $1 \leq i \neq j \leq n$

Examples of Network Data Structure

	1	2	3	4	5
1	0	1	0	0	1
2	1	0	1	1	1
3	0	1	0	1	0
4	0	1	1	0	1
5	1	1	0	1	0

Figure: Adjacency Matrix

Kamada and Kawai: Algorithm

- Before starting the minimization process, the initial positions of particles must be specified
- Experiments have found that the influence of initial positions is trivial on resultant pictures

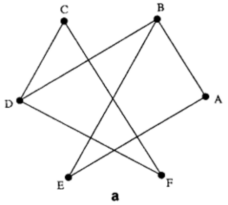
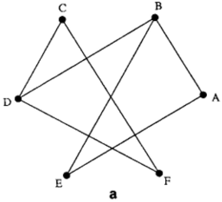


Figure: Regular n-polygon circumscribed by a circle whose diameter is L_0

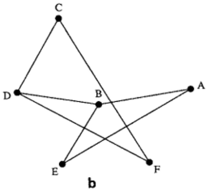
Kamada and Kawai: Algorithm

- initialize p_1, p_2, ..., p_n
- while $(\Delta_m > \epsilon)$
  - let p_m be the particle satisfying $\Delta_m = \max\Delta$
    - compute $\delta_x and \delta_y$
    - $x_m = x_m + \delta x$
    - $y_m = y_m + \delta y$



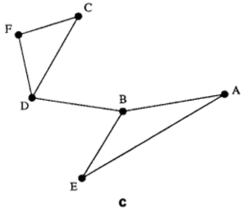
Kamada and Kawai: Algorithm

- while $(\Delta_m > \epsilon)$
  - let p_m be the particle satisfying $\Delta_m = \max\Delta$
    - compute $\delta_x and \delta_y$
    - $x_m = x_m + \delta x$
    - $y_m = y_m + \delta y$



Kamada and Kawai: Algorithm

- while ($\Delta_m > \epsilon$)
  - let p_m be the particle satisfying $\Delta_m = \max \Delta$
    - compute $\delta_x and \delta_y$
    - $x_m = x_m + \delta x$
    - $y_m = y_m + \delta y$



Kamada and Kawai: Algorithm

- while $(\Delta_m > \epsilon)$
  - let p_m be the particle satisfying $\Delta_m = \max\Delta$
    - compute $\delta_x and \delta_y$
    - $x_m = x_m + \delta x$
    - $y_m = y_m + \delta y$

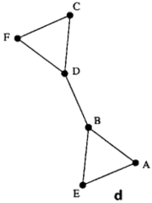
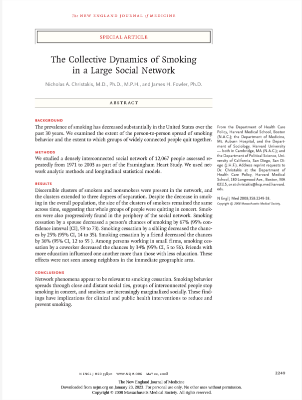


Table of Contents

Application: Smoking Behavior

The Collective Dynamics of Smoking in a Large Social Network

Nicholas A. Christakis and James H. Fowler



The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL ARTICLE

The Collective Dynamics of Smoking in a Large Social Network

Nicholas A. Christakis, M.D., Ph.D., M.P.H., and James H. Fowler, Ph.D.

ABSTRACT

BACKGROUND

The prevalence of smoking has decreased substantially in the United States over the past 30 years. We examined the extent of the person-to-person spread of smoking behavior and the extent to which groups of widely connected people quit together.

METHODS

We studied a densely interconnected social network of 12,067 people assessed repeatedly from 1971 to 2003 as part of the Framingham Heart Study. We used network analytic methods and longitudinal statistical models.

RESULTS

Discernible clusters of smokers and nonsmokers were present in the network, and the clusters extended to three degrees of separation. Despite the decrease in smoking in the overall population, the size of the clusters of smokers remained the same across time, suggesting that whole groups of people were quitting in concert. Smokers were also progressively found in the periphery of the social network. Smoking cessation by a spouse decreased a person's chances of smoking by 67% (95% confidence interval [CI], 59 to 73). Smoking cessation by a sibling decreased the chances by 25% (95% CI, 14 to 35). Smoking cessation by a friend decreased the chances by 36% (95% CI, 12 to 55). Among persons working in small firms, smoking cessation by a coworker decreased the chances by 34% (95% CI, 5 to 56). Friends with more education influenced one another more than those with less education. These effects were not seen among neighbors in the immediate geographic area.

CONCLUSIONS

Network phenomena appear to be relevant to smoking cessation. Smoking behavior spreads through close and distant social ties, groups of interconnected people stop smoking in concert, and smokers are increasingly marginalized socially. These findings have implications for clinical and public health interventions to reduce and prevent smoking.

From the Department of Health Care Policy, Harvard Medical School, Boston (N.A.C.); the Department of Medicine, Mt. Auburn Hospital, and the Department of Sociology, Harvard University — both in Cambridge, MA (N.A.C.); and the Department of Political Science, University of California, San Diego, San Diego (J.H.F.). Address reprint requests to Dr. Christakis at the Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave., Boston, MA 02115, or at christakis@hcp.med.harvard.edu.

N Engl J Med 2008;358:2249-58.

Application: Smoking Behavior

- Framingham Heart Study data from repeated assessments between 1971 to 2003
- 5124 egos from the offspring cohort
- Any persons to whom these subjects are linked, in any of the Framingham Heart Study cohorts, are included as social contacts
- Densely interconnected social network of 12,067 people

Kamadai Kawai Network Visualization

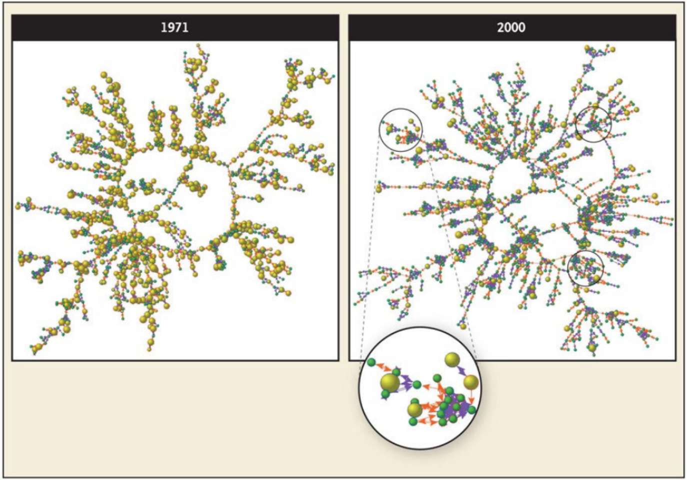


Figure: A random sample of 1000 subjects in the social network from the Framingham Heart Study

Application: Smoking Behavior

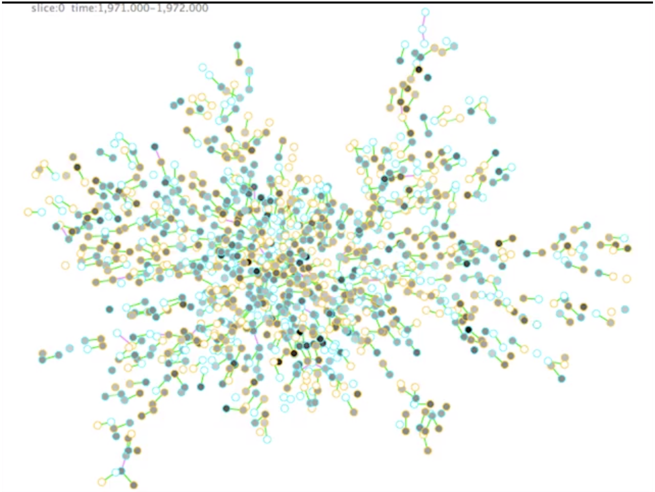
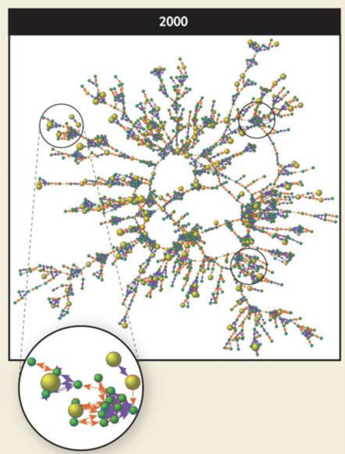


Figure: Dynamic graphic representation by means of the Social Network Image Animator

Application: Smoking Behavior



- By 2000, most people had stopped smoking (prevalence declined from 65.9% to 22.3% over the study period)
- Those who still smoked were more likely to be at the periphery of the network
- Relatively separate clusters of smokers and nonsmokers

Longitudinal Statistical Model

Longitudinal logistic regression model

- Response Variable: Subject's smoking status at a given time (t+1)
- Explanatory Variables: age, sex, education, smoking status at the previous time point (t), and smoking status of his or her contacts at times (t) and (t+1)

Longitudinal Statistical Model

Simplified Design Matrix for 2 subject network

	Age	Sex	Smoke_t	ConSmoke_t	ConSmoke_{t+1}	
Subject A	30	F	1	1	1	Contact 1
	30	F	1	1	0	Contact 2
Subject B	46	M	0	1	1	
	46	M	0	0	1	
	46	M	0	0	0	

Longitudinal Statistical Model

Generalized Estimating Equations

- Semi-parametric model
- Captures the covariance structure of the same subject across examinations over time and across subject-contact pairs with an independent working correlation structure
- Time-lagged dependent variable eliminated serial correlation in the errors

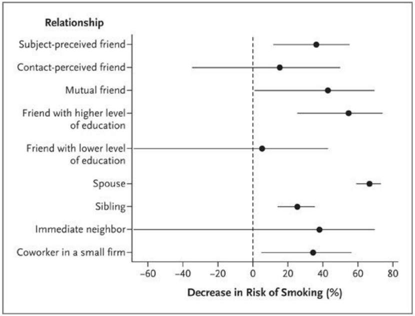
Longitudinal Statistical Model

Table S8: Association of Alter Smoking and Ego Smoking

	Alter Type						
	Ego-Perceived Friend	Mutual Friend	Alter-Perceived Friend	Spouse	Sibling	Immediate Neighbor	Coworker
Alter Currently Smokes	0.51	0.66	0.21	1.19	0.33	0.58	-0.01
	(0.19)	(0.33)	(0.27)	(0.12)	(0.08)	(0.37)	(0.09)
Alter Previously Smoked	-0.53	-0.81	-0.04	-0.47	0.03	-0.56	-0.04
	(0.18)	(0.34)	(0.23)	(0.11)	(0.08)	(0.38)	(0.09)
Ego Previously Smoked	4.51	4.49	5.24	5.26	4.58	6.10	4.28
	(0.21)	(0.38)	(0.33)	(0.14)	(0.14)	(0.57)	(0.40)
Wave 3	0.87	0.81	0.86	1.09	0.91	1.69	0.51
	(0.21)	(0.35)	(0.31)	(0.12)	(0.14)	(0.60)	(0.32)
Wave 4	0.92	0.74	1.81	1.14	0.90	1.61	0.46
	(0.21)	(0.38)	(0.32)	(0.12)	(0.13)	(0.51)	(0.34)
Wave 5	0.68	0.31	1.11	1.17	0.93	1.52	0.61
	(0.22)	(0.40)	(0.30)	(0.14)	(0.14)	(0.58)	(0.37)
Wave 6	0.61	0.44	1.12	1.20	0.95	1.51	0.09
	(0.26)	(0.50)	(0.41)	(0.16)	(0.15)	(0.71)	(0.47)
Wave 7	1.00	0.68	1.18	1.38	1.04	1.90	-0.18
	(0.26)	(0.50)	(0.38)	(0.16)	(0.16)	(0.65)	(0.53)
Ego's Age	-0.03	-0.03	-0.03	-0.03	-0.03	-0.05	0.01
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.02)	(0.01)
Ego Female	0.09	0.13	0.12	0.20	0.18	-0.20	0.61
	(0.13)	(0.25)	(0.19)	(0.08)	(0.09)	(0.31)	(0.24)
Ego's Years of Education	-0.12	-0.10	-0.08	-0.05	-0.11	-0.17	-0.07
	(0.03)	(0.06)	(0.05)	(0.02)	(0.02)	(0.08)	(0.05)
Constant	-1.18	-1.64	-2.73	-3.48	-1.44	-0.84	-3.92
	(0.70)	(1.42)	(1.03)	(0.42)	(0.46)	(1.54)	(1.17)
Deviance	280	85	134	713	1668	68	885
Null Deviance	594	169	322	1667	3742	194	1659
N	3549	1083	2126	10522	21097	1019	8656

Figure: Parameter estimates in the form of beta coefficients

Results



- If a subject stated that a contact was his or her friend, the chance that the subject would be a smoker decreased by 36 % [12-55] if the contact stopped smoking.
- Between mutual friends, the subject's risk of smoking decreased by 43 % [1-69]
- No significant effect of smoking cessation when the friendship was perceived by the contact but not by the subject

Smoking Cessation Conclusions

Additional Findings

- Local smoking-cessation cascades
- Smokers became increasingly marginalized and network become more polarized with respect to smokers and nonsmokers
- Educational background of connected people mattered

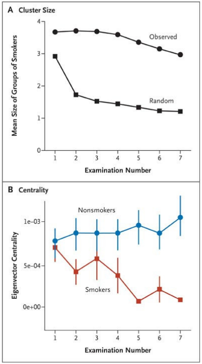


Figure: Cluster Size and the Centrality of Smokers across Time

Smoking Cessation Conclusions

Network phenomena can be exploited to spread positive health behaviors

- Collective interventions may thus be more effective than individual interventions
- Medical and public health interventions to encourage people to quit smoking might be more cost-effective than initially presumed
- Sweeping policy approaches may be usefully supplemented by interventions targeting small groups

Table of Contents

Questions?