Preferential attachment in sexual networks

Birgitte Freiesleben de Blasio[†], Åke Svensson[‡], and Fredrik Liljeros[§¶]

[†]Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, POB 1122, Blindern, N-0317 Oslo, Norway; and Departments of [‡]Mathematics and [§]Sociology, Stockholm University, S-106 91 Stockholm, Sweden

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved April 26, 2007 (received for review December 20, 2006)

Many social networks are characterized by a highly uneven distribution of links. The observed skewed distributions have in several cases been attributed to preferential attachment (PA), a tendency among nodes in a growing network to form new links preferentially to nodes with high numbers of links. We test the PA conjecture in sexual contact networks. A maximum likelihood estimation-based expectation–maximization fitting technique is used to model new partners over a 1-year period based on the number of partners in foregoing periods of 2 years, 4 years, and lifetime. The PA model is modified to account for individual heterogeneity in the inclination to find new partners and fitted to Norwegian survey data on heterosexual men and women. Results show evidence of nonrandom, sublinear PA when comparing the growth in 3- to 5-year periods. The potential implications of these findings are discussed.

sexual behavior | sexually transmitted diseases

In comparison with the observed incidence of sexually transmitted infections in modern societies, the average number of sexual contacts in national populations is surprisingly low. It has been suggested that the endemic and epidemic spread is driven by smaller subsets (core groups) of the population, in which members have significantly higher numbers of partners and a preference for selecting partners within the group (1). Several studies have, however, recently reported a highly skewed distribution of sexual contacts without a clear core group.

The tail of the sex partner distribution is often modeled by a simple power law, that is, the probability mass function (pmf) of sexual partners $P(j)$ have the functional form $P(j) \approx Cj^{-\gamma}$ for some excess $j > j_t$, where C and γ are positive constants. The first suggestion of power law scaling was published by Colgate *et al.* (2) in 1989 from data on homosexual men seen at a sexually transmitted infection clinic in London. More recently, Liljeros *et al.* (3) observed a power law in population data from Sweden, which was later supported by population studies in Burkina Faso (4), Uganda and the United States (5), and Britain and Zimbabwe (6). With one exception (women from the Rakai district in Uganda), the reported slopes are close to $\gamma = 3$; the range for men being $2.8 \leq \gamma \leq 5.4$ and the range for women being $3.0 \leq \gamma \leq 4.2$.

The finding of a power law has been subject to some controversy, and the question has been raised whether a power-law function adequately fits the data. The available sexual data are limited, and the published studies reveal power-law scaling of over one to two orders of magnitude. The limited scaling regime is not sufficient to distinguish a power law from other heavy-tailed distributions, such as log-normal or stretched exponential (Weibull), both of which have characteristic scales and curve away with exponential decay for large enough j. A study by Handcock (7) suggests that a log-normal distribution provides the best description of the data when lower numbers of partners are included.

Here, we address the question of the origin of the skewed distribution by counting sexual partners in overlapping time intervals. The analysis is restricted to one sex (men or women) at a time. This separation is necessary because the partner identities are not known. Presumably, the individuals are not interconnected but have partners outside the study group. A new contact can be regarded as adding a new link to an observed node, i.e., to increase the degree of the node. The separation of men and women has the further advantage that gender-specific differences in reported numbers of partners are not mixed in the estimation procedure.

The aim of our analysis is to answer the following questions: To what extent can we use information on partner numbers to predict future partner numbers, and, if one can, what does it tell us about the distribution of partners in the network? We use a statistical method developed particularly for the present study to analyze cumulative numbers of sex partners in survey data.

The observation that success breeds success is common in many situations. In sociology this dynamic phenomenon is called the Matthew effect (8); in economics, it is called increasing return (9); and, in complex network theory, it is usually referred to as preferential attachment (PA) (10). We will hereafter use the latter term to denote a situation in which the chance of having a new partner increases with the quantity of sexual contacts within a given time frame.

People do not have an equal probability for having sexual contacts. For one thing, people are not perceived as being equally sexually attractive. Second, people have personal preferences regarding emotional involvement with sex partners and promiscuity and different attitudes toward commercial sex. Third, people are affected by their social environment and their religious and normative values. In addition to these more or less static individual properties, there are dynamic social and psychological mechanisms that could encourage a tendency to acquire new partners. For instance, studies have shown a positive correlation between knowledge that a person has many partners and the perceived attractiveness of that person (11). In addition, having new partners can be psychologically addictive (12), and flirting skills are likely to improve with practice, potentially resulting in higher numbers of successful pick-ups.

Power-law distributions and other types of skewed distributions are widespread in social, biological, technical, and information networks. A number of generative network models have been proposed to explain the data (13). Commonly, the models are based on (*i*) constant network growth and (*ii*) preferential linking to nodes with many connections. The term PA was introduced by Barabási and Albert (10) in the context of World Wide Web networks. In their model, new nodes attach links to existing nodes k with a probability proportional to their degree of links,

$$p_a(k) = \frac{j_k}{\sum_i j_i}, \qquad [1]$$

Author contributions: B.F.d.B., Å.S., and F.L. designed research; B.F.d.B., Å.S., and F.L. performed research; B.F.d.B. and Å.S. analyzed data; and B.F.d.B., Å.S., and F.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Abbrevaitions: pmf, probability mass function; PA, preferential attachment.

[¶]To whom correspondence should be addressed. E-mail: fredrik.liljeros@sociology.su.se.

This article contains supporting information online at www.pnas.org/cgi/content/full/0611337104/DC1.

www.pnas.org/cgi/doi/10.1073/pnas.0611337104

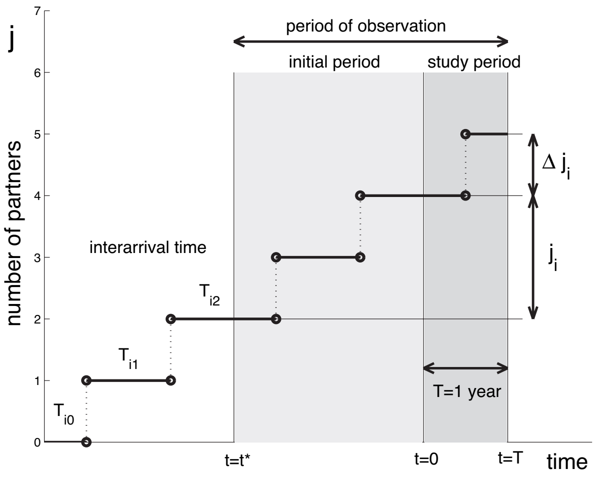

Fig. 1. Schematics of a sample path showing the interarrival (waiting) times T_{i1}, T_{i2}. The observation period consists of two time intervals, the initial period, and the study period. The partner numbers in the first period j_i are used to model the number of new partners Δj_i in the study period (shaded region).

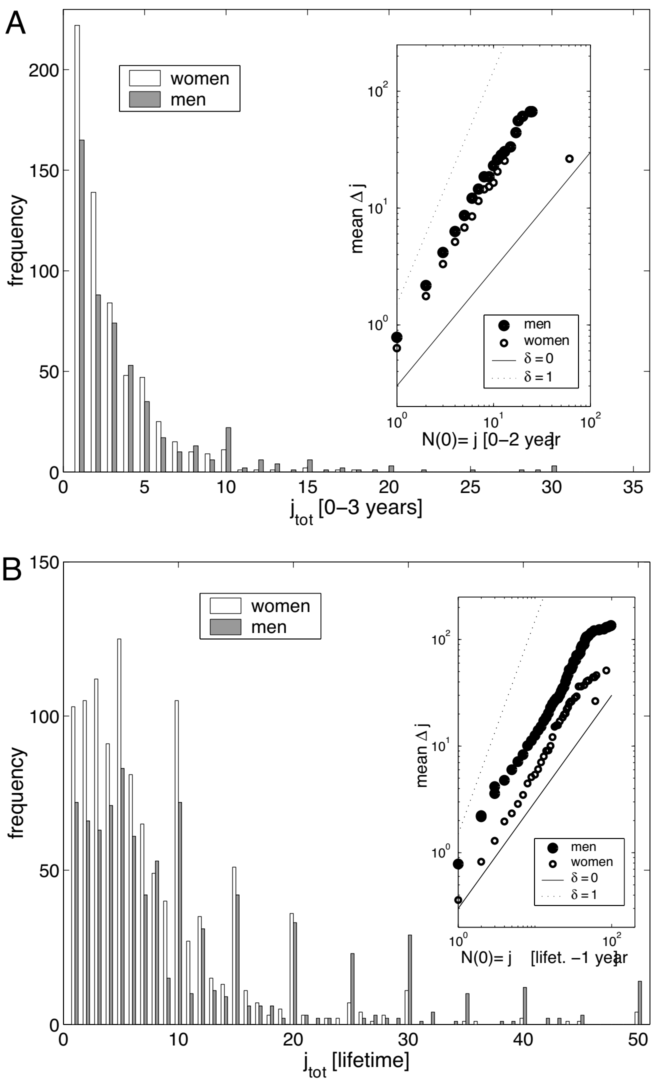

Fig. 2. Histograms showing the distribution of total partners in the observation periods for women (white) and men (shaded) in the Norwegian survey: 3-year period (*A*) and lifetime (*B*). (*A Inset* and *B Inset*) Double-logarithmic plots of the cumulative average numbers of new partners in the 1-year study period for women (circles) and men (stars) plotted as a function of the number of partners *j* in the foregoing period. Thus, the mean values are group averages among all individuals having exactly *j* partners in the initial part of the observation period.

and yields a power-law pmf with $P(j) \sim j^{-3}$. Several modified PA models that take different aspects of the network growth into consideration have been suggested (13).

As early as 1925, Yule (14) published a stochastic preferential growth model to describe the uneven distribution of species among plant genera. The model was later generalized by Simon (15). Adapted to networks, preference in the Yule–Simon process is defined with respect to groups of nodes $[j]$ with identical connectivity j. The probability that a member of group $[j]$ will receive a link is proportional to the abundance of links in the group; that is,

$$p_a([j]) = \frac{n(j)j}{\sum_i n(i)i}, \qquad \textbf{[2]}$$

where $n(j)$ is the number of nodes in the network with degree j. The model generates an asymptotic pmf $P(j) \sim j^{-(1+(1/(1-\alpha)))}$, where α is the ratio of the node versus link creation rates. The two models are closely related, and the Barabási–Albert model may be mapped into a subclass ($\alpha = 1/2$) of the Yule–Simon model (16).

PA in evolving networks is measured by calculating the rate $\pi(j)$ at which groups of nodes $[j]$ with identical connectivity form new links during a small time interval Δt (17). The method has been used to estimate PA in scientific citation and coauthor networks and in author collaboration networks and the Internet (17, 18). The function is described by

$$\pi(j,t) = \frac{\sum_k \Delta j_k}{n(j,t)} \approx C(t)j^{\delta}, \qquad \textbf{[3]}$$

where Δj_k is the number of new links that attach nodes in group j during Δt and $n(j, t)$ is the quantity of nodes with j links at the start of time period t. Eq. **3** holds for short time intervals during which the total number of nodes is roughly constant, $N(t + \Delta t) \approx N(t)$, and for steady-state networks. This latter condition is usually satisfied for linearly expanding networks for which $\langle j \rangle$ is constant.

The dependence of $\pi(j)$ on j is found by plotting the functions. For a linearly growing network, the functional form of the asymptotic pmf can be determined from the δ exponent in Eq. **3** (19, 20). According to the PA hypothesis, $\pi(j)$ should increase monotonically with $j(\delta = 1)$, and linear preference is required for generating fat-tailed pmfs with $P(j) \propto j^{-\gamma}$. For sublinear exponents ($0 \leq \delta < 1$), the pmf is a stretched exponential, $P(j) \propto j^{-\gamma} \exp(-(b(\gamma)/(1-\gamma))j^{1-\gamma})$, where b is a constant depending on γ. The special situation of absent preference ($\delta = 0$) reduces the rate $\pi(j)$ to a constant. In this case, $P(j) \propto \exp(-j)$ is in agreement with the Poisson distribution of a random graph. Finally, for ($\delta > 1$), the growth leads to a gelation-like behavior in which one node is basically connected to all other nodes in the system.

The graphical procedure is hindered by finite-sample effects producing strong fluctuations for large j, and it is insufficient to make an inference about the growth process. Here we present a statistical method that takes these complications into account, and we use it to make an assessment of PA in sexual networks.

Data

The estimation of PA requires ungrouped data on partner numbers during at least two consecutive time intervals. To examine temporal and size-dependent effects, information on sexual contacts in several successive intervals is needed, preferably covering both

Table 1. Maximum likelihood estimation parameter estimates for the PA model

Data		Parameter estimates								Deviance (Dev) statistics		
Study periods	Obs.	α	95% C.I.	β per year	95% C.I.	δ	95% C.I.	ε per year	95% C.I.	Dev	Dev*, mean (SD)	*P* value
Men												
3-year	930	3.50	2.5–4.6	0.27 (0.26)	0.23–0.31	0.62 (0.67)	0.51–0.73	0.76 (0.75)	0.66–0.85	215.2	176.7 (15.4)	0.0062
5-year	919	1.9	1.4–2.6	0.18 (0.17)	0.14–0.23	0.57 (0.61)	0.48–0.66	0.59 (0.58)	0.51–0.70	308.0	278.6 (18.7)	0.0580
Lt	950	9.6	6.4–17.1	0.80 (0.77)	0.64–0.96	0.26 (0.26)	0.19–0.31	0.84 (0.83)	0.74–0.95	1,051.3	319.7 (21.3)	<0.001
Women												
3-year	1,220	2.9	2.0–4.9	0.19 (0.19)	0.16–0.22	0.54 (0.60)	0.41–0.66	0.65 (0.65)	0.57–0.74	92.0	102.8 (12.6)	0.8036
5-year	1,190	1.5	1.2–2.2	0.14 (0.13)	0.11–0.17	0.57 (0.64)	0.47–0.67	0.44 (0.42)	0.38–0.51	136.5	155.7 (15.4)	0.8943
Lt	1,183	0.5	0.4–0.5	0.63 (0.41)	0.38–0.94	0.29 (0.40)	0.20–0.40	0.34 (0.28)	0.26–0.40	356.1	231.3 (21.2)	<0.001

The maximum likelihood estimation parameter estimates for the PA model with random factors $\pi(\hat{\Theta})$ together with the 95% bootstrap confidence intervals (C.I.). The model fit was evaluated by using deviance test statistics (see *SI Text*). Under the null hypothesis, the derived model can generate the observed data. Hence, small *P* values correspond to a lack of fit. For comparison, the estimates derived from the basic preferential model, $\pi(\theta)$, have been added in parentheses for β and ε. Lt, lifetime; Obs., observations.

*Sample of deviance generated from bootstrapping the data.

short and extended time scales. Not many sexual surveys contain partner information at this level of detail.

The National Survey of Sexual Behavior in Norway was conducted by the Norwegian National Institute for Public Health in 2002 (21). To the best of our knowledge, it is the only study in which questions are asked about their exact partner numbers in more than two time periods. The survey is based on 10,000 written questionnaires that were mailed to a random sample of Norwegians between the ages of 18 and 49 years. The respondents supplied information about partner numbers during the previous 1, 3, and 5 years as well as the total number of sexual contacts. The study had a low response rate of 35%, with women being slightly overrepresented. The sample was representative of the Norwegian population with regard to regional, community size, household income, educational level, and occupation (21).

We excluded from the analysis respondents who reported having homosexual contacts (5–10%). Before the analysis, partner numbers during the previous year were adjusted to include new partners only; the procedure involved subtracting one partner from the group of people having a steady partner for >12 months. The data used for the analyses in this paper can be provided upon request.

Model

A general framework for statistical inference for the PA process was developed by Svensson (22). Here we provide a brief description of the model that was used to estimate transition probabilities in the contact networks.

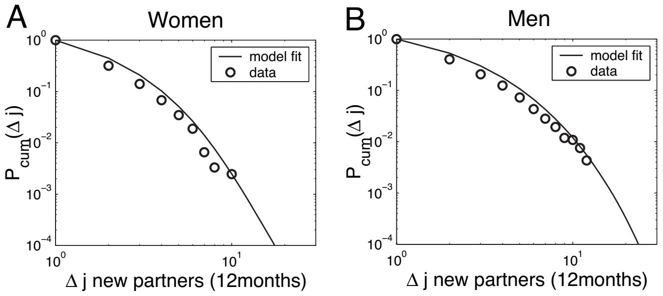

Fig. 3. The cumulative probability of new partners $P_{\text{cum}}(\Delta j) = \Sigma_{i \geq j} P(\Delta i)$ as function of the new partners Δj in the 1-year study period. The probabilities were calculated by using the model parameter (Table 1) and by conditioning on the initial numbers of partners $j_1, j_2, \ldots j_n$ in the study populations. (*A*) Results for women in the 3-year observation period. (*B*) Results for men in the 3-year observation period.

The acquisition of new sexual contacts can be modeled as a pure birth process, with a discrete state space $j = \{0, 1, 2, \ldots\}$, counting the total number of sexual contacts in a person's life, and the transitions $j \to j + 1$ describing the events associated with having a new partner. A random selection of n individuals is observed during two overlapping time periods with a shared end point. We name them the initial period and the study period (see Fig. 1). Our aim is to study how the numbers of new partners increase during the study period depending on each person's individual history of new partners during the initial period. In the following analysis, we choose different initial periods, but in all cases the study period is the last single year covered by the survey. The vector $\mathbf{N} = N_1(\cdot), N_2(\cdot), \ldots N_n(\cdot)$ counts the numbers of new partners during this interval. For convenience, we set $t = t^*$ at the start of the period of observation, $t = 0$ at the start of the study period and $t = T$ (= 1 year) at the end of the observation period. With this notation, $N_i(0) = j_i$ is the number of new partners for individual i during the initial period and Δj_i is the number of new partners during the study period. Thus, $N_i(T) = j_i + \Delta j_i$ is the total number of new partners during the entire observation period. We will set up a model that describes the distribution of the random variable Δj_i given j_i.

Let π be an intensity vector describing the rate of transitions between the different states,

$$\pi = (\pi_0, \pi_1, \ldots) \qquad [4]$$

where each term π_j for $j = \{0, 1, 2, \ldots\}$ is the one-step probability per year that a person with exactly j partners will acquire a new partner. In accordance with the PA scenario, Eq. **1**, the following parametric model $\pi(\theta)$ is assumed.

$$\pi_j(\theta) = \begin{cases} \beta & j = 0 \\ \varepsilon j^{\delta} & j \geq 1 \end{cases}. \qquad [5]$$

The process of having a first partner is considered here separately.

In this model, the jump intensities for all individuals are assumed to be equal. One may expect considerable individual variation in partner numbers depending, for example, on socioeconomic factors. Lacking information on such auxiliary variables, heterogeneity is introduced into the model by modifying the intensity vector with a random proportionality factor κ_i for each i individual, $\pi \to \kappa_i \pi = (\kappa_i \pi_0, \kappa_i \pi_1, \ldots)$. The frailty terms are drawn independently from a gamma distribution,

$$g_\alpha(\kappa) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \kappa^{\alpha - 1} \exp(-\alpha \kappa) \qquad [6]$$

Table 2. Frailty model parameter estimates for the stratified data sets during 3- and 5-year observation periods

Data		Parameter estimates									Deviance (Dev) statistics		
Study periods and groups	Obs.	α	95% C.I.	β per year	95% C.I.	δ	95% C.I.	ε per year	95% C.I.	Dev	Dev,* mean (SD)	*P* value
Men												
3-year												
18–29 y	270	5.7	2.0–6.0	0.83	0.75–1.70	0.52	0.35–0.61	0.95	0.64–1.02	129.9	120.7 (12.1)	0.2246
30–39 y	328	1.5	1.1–3.0	0.26	0.19–0.33	0.66	0.50–0.85	0.63	0.48–0.81	127.4	123.7 (13.5)	0.3906
40–49 y	329	4.5	2.0–16.1	0.13	0.09–1.17	0.80	0.56–0.95	0.60	0.46–0.78	78.8	73.9 (9.8)	0.3071
Single	301	11.5	7.0–27.0	1.26	1.00–1.58	0.50	0.40–0.59	1.24	1.08–1.43	172.3	141.8 (13.1)	0.0101
Cohab	625	1.2	0.8–2.6	0.14	0.11–0.17	0.67	0.42–0.87	0.35	0.27–0.45	98.5	92.3 (12.1)	0.3054
5-year												
18–29 y	260	3.4	2.0–5.5	1.13	0.74–1.72	0.48	0.35–0.60	0.77	0.64–0.94	181.4	165.0 (13.8)	0.1189
30–39 y	328	1.3	0.9–2.1	0.13	0.08–0.20	0.63	0.48–0.77	0.49	0.37–0.63	192.5	179.2 (15.5)	0.1965
40–49 y	328	1.8	1.1–3.1	0.08	0.05–0.12	0.65	0.44–0.83	0.47	0.34–0.63	111.6	103.1 (11.7)	0.2335
Single	292	7.6	5.0–13.0	1.26	0.90–1.77	0.46	0.37–0.55	1.10	0.93–1.28	224.9	207.4 (14.9)	0.1191
Cohab	623	0.9	0.7–1.6	0.10	0.07–0.13	0.63	0.43–0.78	0.26	0.20–0.34	145.3	137.5 (14.9)	0.2993
Women												
3-year												
18–29 y	434	4.5	2.5–16.2	0.46	0.37–0.57	0.64	0.51–0.79	0.62	0.52–0.72	76.5	87.1 (11.2)	0.8270
30–39 y	426	1.6	1.6–3.5	0.12	0.09–0.16	0.36	0.13–0.62	0.71	0.55–0.89	59.8	60.8 (10.1)	0.5401
40–49 y	357	3.3	1.6–17.0	0.12	0.08–0.16	0.50	0.20–0.77	0.62	0.47–0.82	32.3	39.5 (8.1)	0.8105
Single	361	5.5	3.1–13.2	1.33	1.05–1.74	0.32	0.19–0.46	1.14	0.98–1.32	86.8	84.6 (11.2)	0.4215
Cohab	850	5.5	2.0–15.5	0.09	0.07–0.11	0.56	0.34–0.84	0.30	0.24–0.35	51.9	39.4 (8.0)	0.0574
5-year												
18–29 y	422	1.8	1.3–3.0	0.64	0.45–0.86	0.60	0.49–0.73	0.45	0.35–0.55	112.6	131.8 (13.9)	0.9164
30–39 y	416	1.5	0.9–2.8	0.07	0.04–0.11	0.57	0.31–0.80	0.42	0.30–0.59	67.2	79.4 (11.3)	0.8601
40–49 y	349	1.2	0.7–3.6	0.08	0.05–0.11	0.36	0.09–0.69	0.50	0.35–0.71	56.7	52.4 (9.5)	0.3232
Single	346	4.5	2.6–8.9	1.96	1.32–3.02	0.38	0.25–0.50	0.92	0.77–1.10	125.1	126.2 (13.4)	0.0082
Cohab	835	2.1	1.2–15.5	0.06	0.04–0.08	0.60	0.40–0.80	0.19	0.14–0.24	75.8	63.7 (10.2)	0.1175

Analysis of data sets stratified by civil status and age. The data are separated into classes of single individuals living alone and those cohabitating (cohab) during the previous 1-year time period before the study ended. The age-stratified data sets were constructed by grouping the sample population into three 10-year age cohorts based on age by the end of the survey. The δ exponents are found to be distinctly sublinear. The parameter estimates β and ε are significantly increased for persons living singly and among young men between the ages of 18 and 29, whereas these baseline parameters are reduced for people in a live-in relationship. Lt, lifetime; Obs., observations.

*Sample of deviance generated from bootstrapping the data.

with mean $\bar{g}_{\alpha}(\kappa) = 1$ and variance $1/\alpha$. In this model the intensities depend on four parameters, $\theta = (\alpha, \beta, \delta, \varepsilon)$.

A detailed description of the model is provided in the supporting information (SI) *Text*.

Results

Explorative Analysis. The mean age for men and women in the study groups was 34 and 33 years with an overall uniform age distribution. Approximately two-thirds of the individuals were cohabiting during the previous 1 year, and one-third of the men and women were living singly.

The range of initial states $N_i(0)$ covers on the order of two decades with the main body of observations centered in the lower region. Only 1% of the women and 3% of the men reported >10 partners during the primary 2 years during the 3-year study, whereas the corresponding values for the primary 4 years during the 5-year study were 2% women and 7% men. In the lifetime study, 26% of the women and 29% of the men reported having >10 partners when excluding the previous year.

Fig. 2 depicts the pmf of final states for two of the observation periods: the 3-year period (Fig. 2*A*) and the lifetime period (Fig. 2*B*) (the 5-year period is shown in SI Fig. 5). Some clustering of final states (10, 15, 20, etc.) is observed. The overrepresentation of rounded values indicates that the data suffers from inaccurate recall, and partner numbers $N_i(T) \geq 10$ should be interpreted as being approximate.

Fig. 2 *A Inset* and *B Inset* provide a graphical estimation of PA in the networks. We show the cumulative rate $\pi(j)_{\text{cum}} = \Sigma_{i \leq j}\, \pi(i)$ to reduce the fluctuations for large j caused by the poor statistics in that region. In this case, the scaling behavior $\pi(j) \sim j^{\delta}$ is replaced by $\pi(j)_{\text{cum}} \sim j^{\delta+1}$ (17). We plotted the mean numbers of new partners $\langle \Delta j \rangle_{\text{cum}} = \Sigma_{i \leq j}\, \Delta i/n(i, t)$ during the last 1 year as functions of quantities of partners j during the foregoing periods $t = 2$ years and lifetime on logarithmic scales. Two lines have been added to assist in the interpretation of the graphs. The expected slope in absence of preference ($\delta = 0$) is shown by a solid line. The case of linear preference ($\delta = 1$) is shown by a dashed line. At first glance, all of the networks seem to fall between these categories.

The short-time longitudinal data (Fig. 2*A*) displays a linear regime, which extends to around $j \approx 10$, whereas the lifetime network data (Fig. 2*B*) appears more complex in that the curves follow an approximately straight line up to around $j \approx 20$–30, followed by a region with possibly higher slopes.

Maximum Likelihood Estimates. Table 1 shows the parameter estimates for the PA model. The test statistics of the basic PA model was not satisfactory. It was therefore ruled out as a candidate model. In addition, no adequate fit to lifetime partner data was obtained. In the following discussion, we address only the studies with short time periods.

The δ exponent is conferred to values between 0.5 and 0.7. In all cases, the bootstrap confidence intervals show sublinear PA with values of >0 and <1. The ε parameter decreases with the length of the initial period. This finding is expected given the constant δ parameter. We describe the same outcome, namely the observed partners during the 1-year study period, regardless of the length of the initial period. In a longer prestudy period, the partner numbers we condition on are larger. Thus, if δ is constant, ε should decrease with time.

The parameter β is estimated from observations of persons with no (new) partners during the prestudy period. This population is likely to depend on the length of the initial period. For a short initial period, the group may consist of persons with a stable relationship, sexually inactive individuals and young persons with no sexual experience. For longer prestudy periods, the population primarily consists of sexually inexperienced individuals.

Women report significantly fewer partners compared with men (compare with Fig. 2); this gender inequality is reflected in a considerable reduction in the baseline β,ε parameters for females.

One way to visualize the model fit is to calculate the expected numbers of new partners in the 1-year study period conditional on the observed distribution of initial states. These values may then be compared with the observed distribution of new partners. Fig. 3 and and SI Fig. 6 show the cumulative pmf $P_{\mathrm{cum}}(\Delta j) = \Sigma_{i \geq j} P(\Delta i)$ plotted as a function of the quantity of new partners during the study period Δj on logarithmic axes for the model (solid line) and the data (symbols). As seen in Fig. 3 and SI Fig. 6, the model provides a good estimate of the probability distribution of contacts in the test interval.

Stratified Analyses. A similar analysis was conducted on data sets stratified by civil status and age (Table 2). We separated the data into classes of single individuals living alone and those cohabitating during the previous 1-year time period before the study ended. The cohabitating status was preferred over marriage status because ≈25% (Statistics Norway) of all live-in relationships in Norway are not registered. The age-stratified data sets were constructed by grouping the sample population into three 10-year age cohorts based on age by the end of the survey. Analyses similar to that shown in Table 2 were conducted on stratified lifetime partner data; however, the test statistics were not acceptable (data not shown). The δ exponents are found to be distinctly sublinear. The parameter estimates β and ε are significantly increased for persons living singly and among young men between the ages of 18 and 29, whereas these baseline parameters are reduced for people in a live-in relationship.

Fig. 4 and SI Fig. 7 present the cumulative partner distribution pmf P_{cum} $(\Delta j) = \Sigma_{i \geq j}$ $P(\Delta i)$ plotted as function of new partners during Δj. In this case, we calculated the expected numbers of new partners separately for each stratum conditional on the initial states. Then the model expectation was found by summing the partner contributions of each group. The model expectations derived from stratification by civil status (cohab model) and 10-year cohorts (age model) are shown in line plots.

As Fig. 4 and SI Fig. 7 show, the modeled distributions of the data stratified by civil status and age are similar for low partner numbers. However, the age-stratified models predict more prolonged tails, which is more similar in shape to the unstratified models (Fig. 3).

Discussion

We have quantified the importance of PA for the formation of new sexual contacts. For this purpose, a framework for statistical inference of a generalized PA mechanism was presented, allowing for random heterogeneity in the subjects. The method was applied to model the distribution of new sex partners during a 1-year period using Norwegian survey data.

Individual heterogeneity in the inclination for making sexual contacts was found essential, as no satisfactory model fit was obtained for the pure PA model. Instead, the PA frailty model produced adequate fit to data when conditioning on contact numbers over the past 2–4 years. Intrinsic growth rates have also been considered by Barabási and Albert (23); heterogeneity was introduced to avoid the correlation between age and connectivity, which otherwise arises naturally in their PA model.

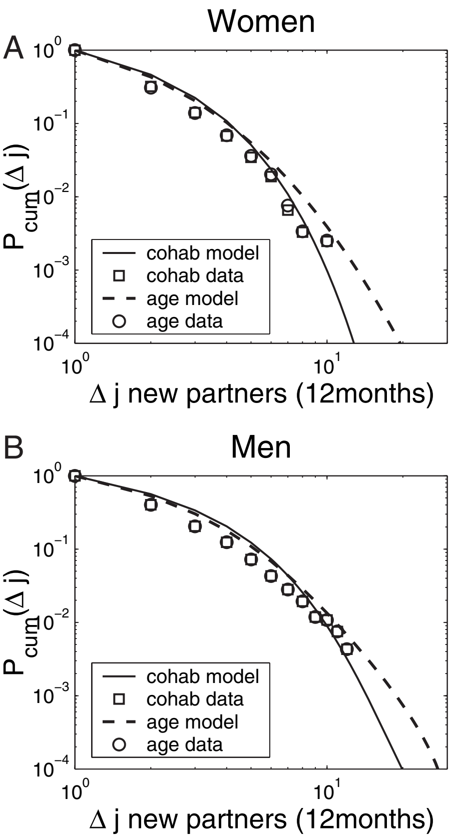

Fig. 4. The cumulative probability of new partners $P_{\mathrm{cum}}(\Delta j) = \Sigma_{i \geq j} P(\Delta i)$ as a function of the new partners Δj in the 1-year study period. The probabilities are calculated by using the model parameter (Table 2). The individuals are partitioned based on their age or civil status, and the expected numbers of new partners in each strata were calculated separately based on the initial numbers of partners $j_1, j_2, \ldots j_n$. These numbers were then summed to provide the pmf of the entire population. (*A*) Results for women in the 3-year observation period. (*B*) Results for men in the 3-year observation period.

The parameter estimates of the two candidate models studied here were in agreement, and the much simpler basic model may be used to gauge their values. This finding is expected because the two models give identical descriptions of the mean intensities in the population. The major difference between the two models is that inclusion of individual growth rates gives frailty models wider confidence intervals for the estimated parameters.

The most salient finding is a scaling-exponent $\delta \sim 0.5$–0.6 with 95% confidence intervals $0 < \delta < 1$. Interestingly, this finding also applies to data stratified by age and cohabitation status. The result is generically of a density mass function of contact numbers belonging to the stretched exponential family. This finding is interesting because it suggests that the reported power-law degree distribution does not emerge from a simple PA process.

However, some caution must be exercised. First, the data material analyzed is limited, and the Norwegian setting may be different from other countries. Second, linear PA growth on the entire graph is not necessary for a power-law network to emerge. In ref. 24, the authors report power-law scaling in the degree distribution of a scientific citation network, with sublinear PA on new nodes, whereas links between existing nodes are linear by degree. In this case, the dynamics are governed by the internal attachment process between old nodes, producing the power-law degree distribution. Unfortunately, because of the lack of partner identities, this particular aspect cannot be studied with the present data.

In the majority of cases, the deviance test statistics gave P values in the range of 0.5–0.8 for women, and 0.05–0.3 for men. Given the highly simplistic nature of the model, these values may be considered satisfactory. Thus, the analyses show evidence of sublinear PA in growing sexual networks.

Young people between 18 and 29 years of age and persons living singly reported higher partner numbers compared with values in the ungrouped data. In Norway, there has been a trend toward a decrease in the age of first sexual encounter during the past 10–15 years. During the same time period, the numbers of lifetime heterosexual partners, particularly among youngsters, have increased (25); similar findings have been reported in Britain (26, 27). The stratified analyses gave consistent estimates for the exponent δ, which were also in agreement with the δ values of the ungrouped data. By contrast, the β and ε parameters, which describe the contact rates among individuals with 0 or 1 prior partners, were often distinct. The interpretation of these results is that contact numbers of people with few partners initially are indicative of general behavior in the strata. Although the parameter confidence intervals were quite wide, the model proved sensitive enough to distinguish both high- and low-value groups.

The finding of insufficient model fit to lifetime partner data is important and reveals the limitations of the PA model. PA is not a mechanism that stands alone; it works within a sociodemographic context. The tendency of an individual to engage in multiple sexual contacts varies with steady partnership and age, in addition to other aspects not addressed here, such as changing social norms and residency, among others. All of these factors causes the model assumption of increasing contact rates with past partner numbers to break down over the long term.

It should be noted that the interpretation of lifetime partner data are not straightforward as we compare individuals by their total number of partners up to some given point in time. Because of this "right censoring," the years of active sexual life among subjects vary substantially. Young people tend to report the fewest partners while simultaneously having high partner change rates. This behavior is inconsistent with the PA scenario and may contribute to the lack of fit to lifetime data. However, no satisfactory model fit was obtained from the age-stratified data (data not shown). This result in turn suggests that the PA scenario is not adequate to model distributions of new partners when conditioning on long-time partner data.

The ability to generalize the present study is limited by several factors. Above all, the response rate of 35% is quite low and is lower than reported values for other European national sexual surveys. This low response rate naturally raises the concern that respondents may differ in some systematic way from those who refused participation; for example, women with extremely high numbers of partners may be underrepresented in national surveys (28). However, we note that the present analysis focuses on the speed at which new partners are obtained. Thus, a low response rate will bias our results only if the responders differ from the general population in their propensity to have new partners during the 1-year study period. It is not necessary that the survey reflect the contact numbers during the initial period in a correct way, because our analyses are made conditional on these numbers.

Studies of sexual behavior are based on self-reports and are exposed to various forms of recollection bias, principally recall difficulties and self-disclosure bias, that is, deliberate misinformation about true behavior, as well as other methodological problems (29). Recall errors may increase with the length of the recall period (30), and contacts seem to be remembered more easily by people with infrequent partner changes (31). The uncertainty of partner numbers was discernible in a clear preference for rounded numbers among persons with many partners. We tested the effect of data clustering by fitting the model to smoothed data, replacing initial and final values $N_i(0)$, $N_i(T)$ with whole numbers within a range $\pm 10\%$ of the original values, and by assuming a flat distribution curve. The procedure improved the model fit substantially for men in the 2- to 4-year studies. The P values were increased by a factor of 10, thus signifying that data clustering impinges on model performance. The procedure did not noticeably affect the parameter estimates. Among women, the effect of data smoothing was negligible, as partner numbers ≥ 10 were reported only rarely. However, the results are highly preliminary. The crude supposition of symmetric variation around the reported values contradicts the gender-specific finding that men and women tend to over- and underreport their partner numbers (32). In addition, recall bias also has been found among individuals reporting one to two partners during a 5-year period (33). Further studies in this area are clearly merited.

When analyzing data stratified by live-in partnerships, no distinction was made as to the cohabitation status of individuals before the 1-year study period. Although the length of the last steady partnership was known, other live-in relationships could not be controlled for. This inconsistency could explain the considerable variation in model fit for these strata.

One effect of PA is that the variation in partner turnover rate in a cohort will increase over time. The current variation in the risk for sexually transmitted infection may therefore partly be a function of the individuals' earlier sexual history instead of sociodemographic differences. PA may, for example, explain the lack of sociodemographic predictors (except age) for genital chlamydia (34). Thus, strategic intervention programs that focus on traditional risk groups may be less effective because a substantial group of individuals with a high number of partners might not be targeted.

We thank Hein Stigum at the Norwegian Institute of Public Health for allowing access to the sexual survey data. This work was supported by the National Institute of Public Health of Sweden, the National Board for Health and Welfare of Sweden, and New and Emerging Science and Technologies (NEST) Project DYSONET 012911 from the Sixth Framework Programme of the European Union.

1. Thomas JC, Tucker MJ (1996) *J Infect Dis* 174(Suppl 2):S134–S143.
2. Colgate S, Stanley E, Hymann J, Layne S, Qualls C (1989) *Proc Natl Acad Sci USA* 86:4793–4797.
3. Liljeros F, Edling C, Amaral L, Stanley E, Åberg Y (2001) *Nature* 411:907–908.
4. Latora V, Marchiori M, Nyamba A, Musumeci S (2006) *J Med Virol* 78:724–729.
5. Jones JH, Handcock MS (2002) *Proc R Soc London Ser B* 270:1123–1128.
6. Schneeberger A, Mercer CH, Gregson SA, Ferguson NM, Nyamukapa CA, Anderson RM, Johnson AM, Garnett GP (2004) *Sex Transm Dis* 31:380–387.
7. Handcock MS, Jones JH (2004) *Theor Popul Biol* 65:413–422.
8. Merdoc R (1968) *Science* 159:56–63.
9. Arthur B (1994) *Increasing Returns and Path Dependence in the Economy (Economics, Cognition, and Society)* (Univ of Michigan Press, Ann Arbor, MI).
10. Barabási AL, Albert R (1999) *Science* 286:509–512.
11. Dugatkin L (2001) *The Imitation Factor: Evolution Beyond The Gene* (Free, New York).
12. Waska R (2006) *Am J Psychoanal* 66:43–62.
13. Dorogovtsev S, Mendes J (2003) *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Univ Press, New York).
14. Yule G (1925) *Phil Trans R Soc London Ser B* 213:21–87.
15. Simon H (1955) *Biometrika* 42:425–440.
16. Bornholdt S, Ebel H (2001) *Phys Rev E* 64:035104.
17. Jeong H, Néda Z, Barabási AL (2001) *Euro Phys Lett* 61:567–572.
18. Newman MEJ (2001) *Phys Rev E* 64:025102.
19. Krapivsky PL, Redner S, Leyvraz F (2000) *Phys Rev Lett* 85:4629–4632.
20. Dorogovtsev SN, Mendes JFF, Samukhin AN (2000) *Phys Rev Lett* 85:4633–4636.
21. Træen B, Stigum H, Magnus P (2003) *Rapport fra Seksualvaneundersøkelsene* (Norwegian Inst of Publ Health, Oslo), available at www.fhi.no/dav/A7A54173BE.pdf. Accessed May 23, 2007.
22. Svensson Å (2005) *Estimating Transition Intensities in Pure Birth Processes Sampled in Time* (Stockholm University, Stockholm), available at www.math.su.se/matstat/reports/seriea/2005/rep11/report.pdf. Accessed May 23, 2007.
23. Barabási AL, Albert R, Jeong H, Bianconi G (2000) *Science* 287:2115a.
24. Barabási AL, Neda Z, Ravasz E, Schubert A, Vicsek T (2002) *Phys A* 311:590–614.
25. Pedersen W, Samuelsen S (2003) *Tidsskr Nor Leageforen* 123:3006–3009.
26. Johnson AM, Mercer CH, Copas AJ, McManus S, Wellings K, Fenton KA, Korovessis C, Macdowall W, Nanchahal K, Purdon S, *et al.* (2001) *Lancet* 358:1835–1842.
27. Wellings K, Nanchahal K, Macdowall W, McManus S, Erens B, Mercer CH, Johnson AM, Copas AJ, Korovessis C, Fenton KA, *et al.* (2001) *Lancet* 358:1843–1850.
28. Brewer DD, Potterat JJ, Garrett SB, Muth SQ, Roberts JM, Kasprzyk D, Montano DE, Darrow WW (2000) *Proc Natl Acad Sci USA* 97:12385–12388.
29. Catania J, Gibson D, Chitwood D, Coates T (1990) *Psychol Bull* 108:339–362.
30. Graham CA, Catania JA, Brand R, Duong T, Canchola JA (2003) *J Sex Res* 40(4):325–332.
31. Kupek E (2002) *BMC Med Res Methodol* 2:14.
32. Wiederman MW (1997) *J Sex Res* 34:375–386.
33. Riska A, Diev V, Smirni E (2002) *ACM SIGMETRICS Perform Eval Rev* 30:6–8.
34. Kučinsklenė V, Šutaltė I, Valiukevičienė S, Milašauskienė Ž, Domeika M (2006) *Medicina* 42(10):885–894.
35. Taylor HM, Karlin S (1998) in *An Introduction To Stochastic Modeling* (Academic, San Diego), pp 333–417.