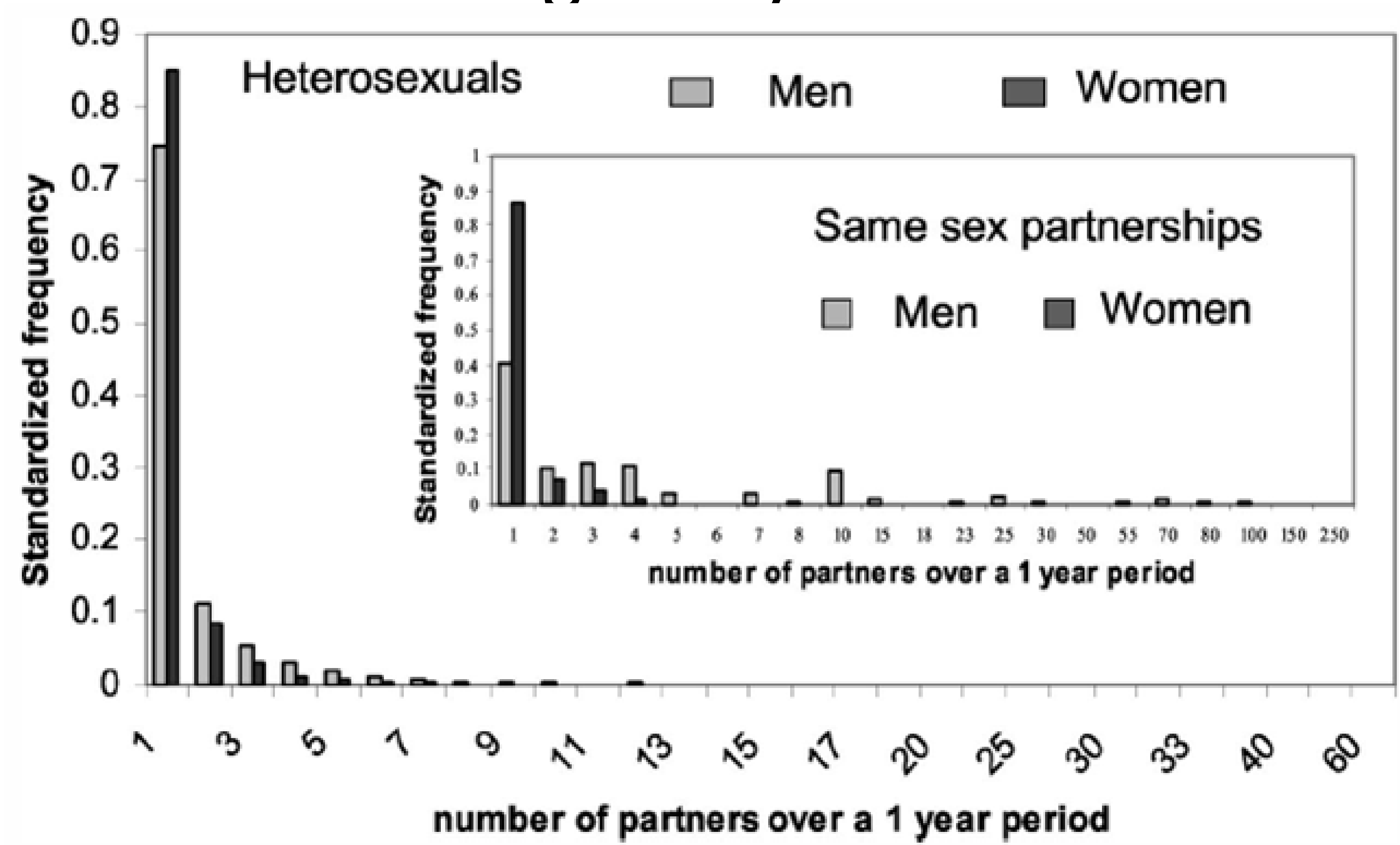
Scale-Free Networks

Nathaniel Osgood

CMPT 858

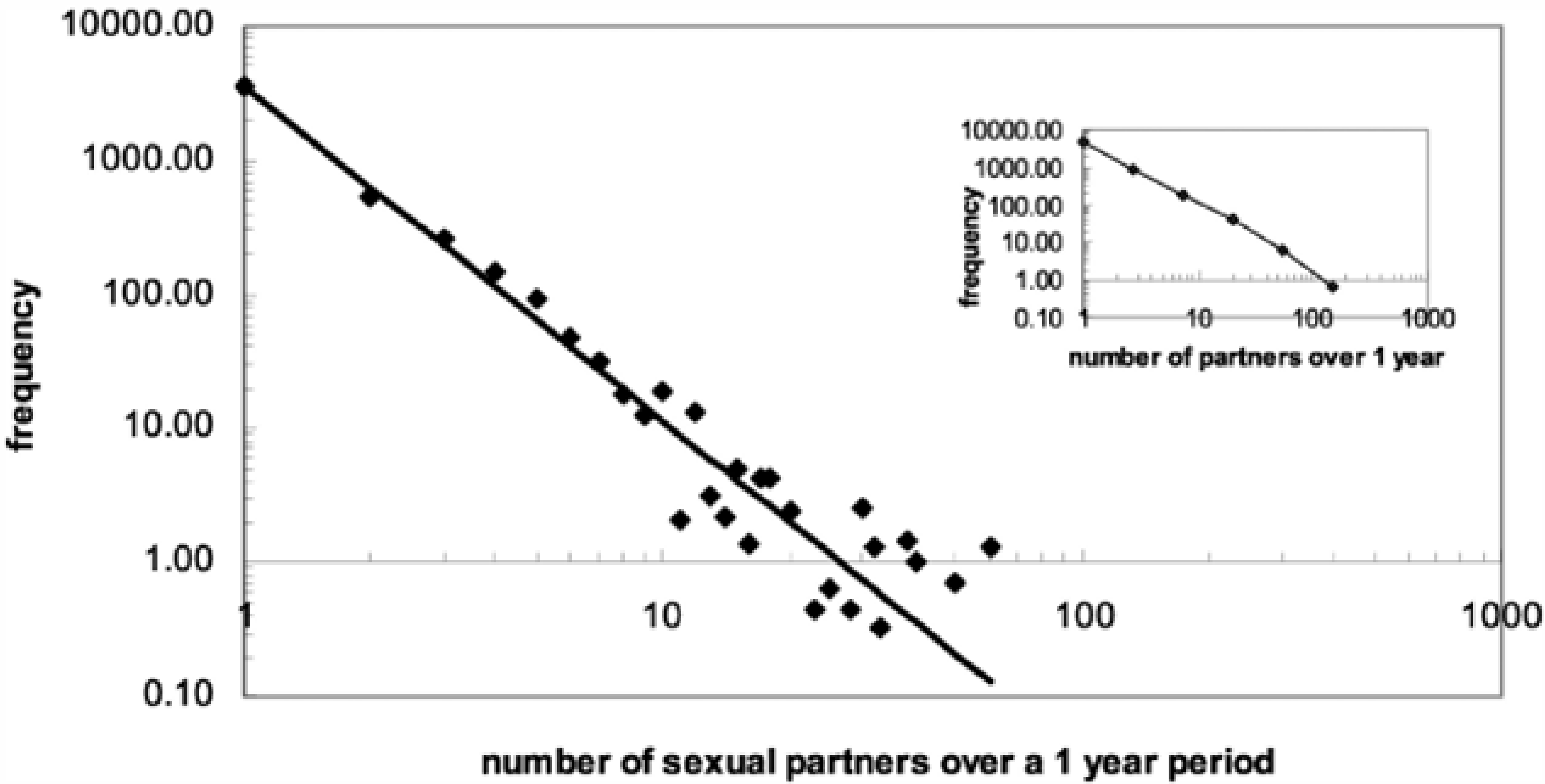
March 7, 2013

Recall: Heterogeneity in Contact Rates



Schneeberger et al., Scale-Free Networks and Sexually Transmitted Diseases:
A Description of Observed Patterns of Sexual Contacts in Britain and Zimbabwe
, Sexually Transmitted Diseases, June 2004, Volume 31, Issue 6, pp 380-387

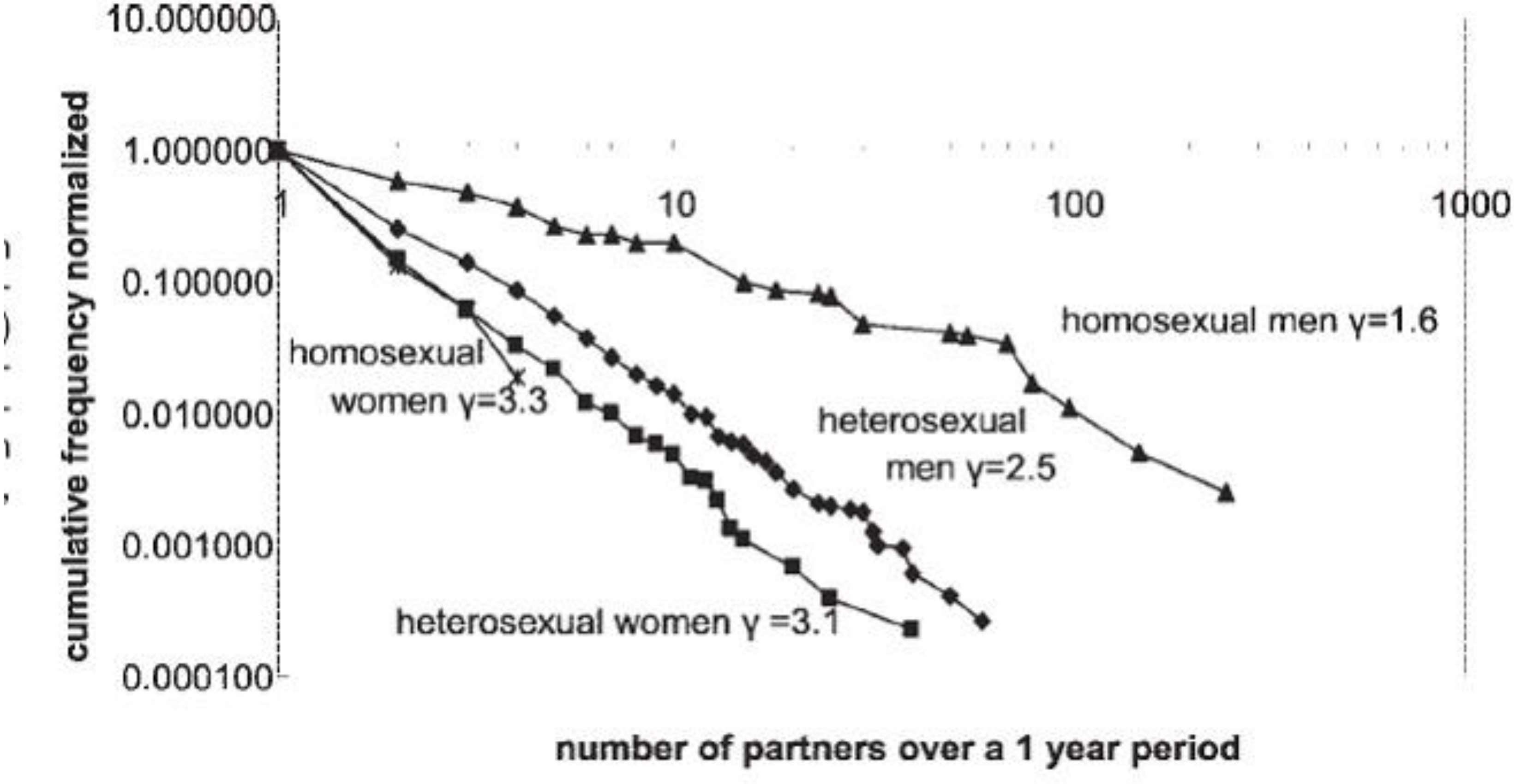
Associated Log-Log Graph



Schneeberger et al., Scale-Free Networks and Sexually Transmitted Diseases:
A Description of Observed Patterns of Sexual Contacts in Britain and Zimbabwe
, Sexually Transmitted Diseases, June 2004, Volume 31, Issue 6, pp 380-387

Heterogeneity in Contact Rates

This may significantly affect the spread of infection in the population!



Schneeberger et al., Scale-Free Networks and Sexually Transmitted Diseases: A Description of Observed Patterns of Sexual Contacts in Britain and Zimbabwe , Sexually Transmitted Diseases, June 2004, Volume 31, Issue 6, pp 380-387

Intuitive Plausibility of Importance of Heterogeneity

- Someone with high # of partners is both
  - More likely to be infected by a partners
  - More likely to pass on the infection to another person
- Via targeted interventions on high contact people, may be able to achieve great “bang for the buck”
- We may see very different infection rates in high contact-rate individuals

- **How to modify classic equations to account for heterogeneity? How affects infection spread?**

Recall: Classic Infection Term

$$\dot{Y} = c\left(\frac{Y}{N}\right)\beta X - \frac{Y}{D}$$

- Xs are susceptibles, Ys are infectives
- c is contacts per unit time
- β is chance a given contact between an infective and a susceptible will transmit infection

Key Step: Disaggregate by Contact Rate

- We break the population up in to groups according to their rate of contacts
- x_i and y_i are susceptibles, infectives who contact i other people per unit time
  - X is divided into x_0, x_1, …
  - Y is divided into y_0, y_1, …

First Attempt

This rate of contact used to be a single constant (c), but now we've captured the Heterogeneity in rates!

$$\dot{y}_i = i\left(\frac{\sum_{j=1}^{\infty} y_j}{N}\right)\beta x_i - \frac{y_i}{D}$$

This is the total number of Infected people

- Here we are capturing the higher levels of risk for someone of activity class *i* as *i* increases (due to higher contact rates)
- Problem:
  - We are assuming that our *i* contacts are equally spread among other people – in fact, they are skewed towards *others* with a high # of contacts!
  - People with high #s of contacts are more likely to be infected

Revised Formulation

This is the total number of contacts per unit time made by infectives!

$$\dot{y}_i = i\left(\frac{\sum_{j=1}^{\infty} j y_j}{\sum_{j=1}^{\infty} j N_j}\right)\beta x_i - \frac{y_j}{D}$$

This is the total number of *contacts per unit time* made by the entire population.

- x_i and y_i are susceptibles, infectives who contact *i* other people per unit time
- The fraction indicates fraction of *contacts in the population* that are with an infective person
  - *i* times this is the rate of contacts with infectives per unit time experienced by a susceptible in class *i*

Force of Infection

$$\lambda = \beta \left(\frac{\sum_{j=1}^{\infty} j y_j}{\sum_{j=1}^{\infty} j N_j} \right)$$

λ will only grow if y grows!

Reformulated Equation

$$\dot{\lambda} = \lambda\left(\beta\frac{E\left[j^2\right]}{E\left[j\right]} - \frac{1}{D}\right)$$

- This is exactly like the normal SIR system, with X = 1,c= $\frac{E\left[j^2\right]}{E\left[j\right]}$
- R_0 is $$\beta\frac{E\left[j^2\right]}{E\left[j\right]}D$$

Reformulating in More Familiar Terms

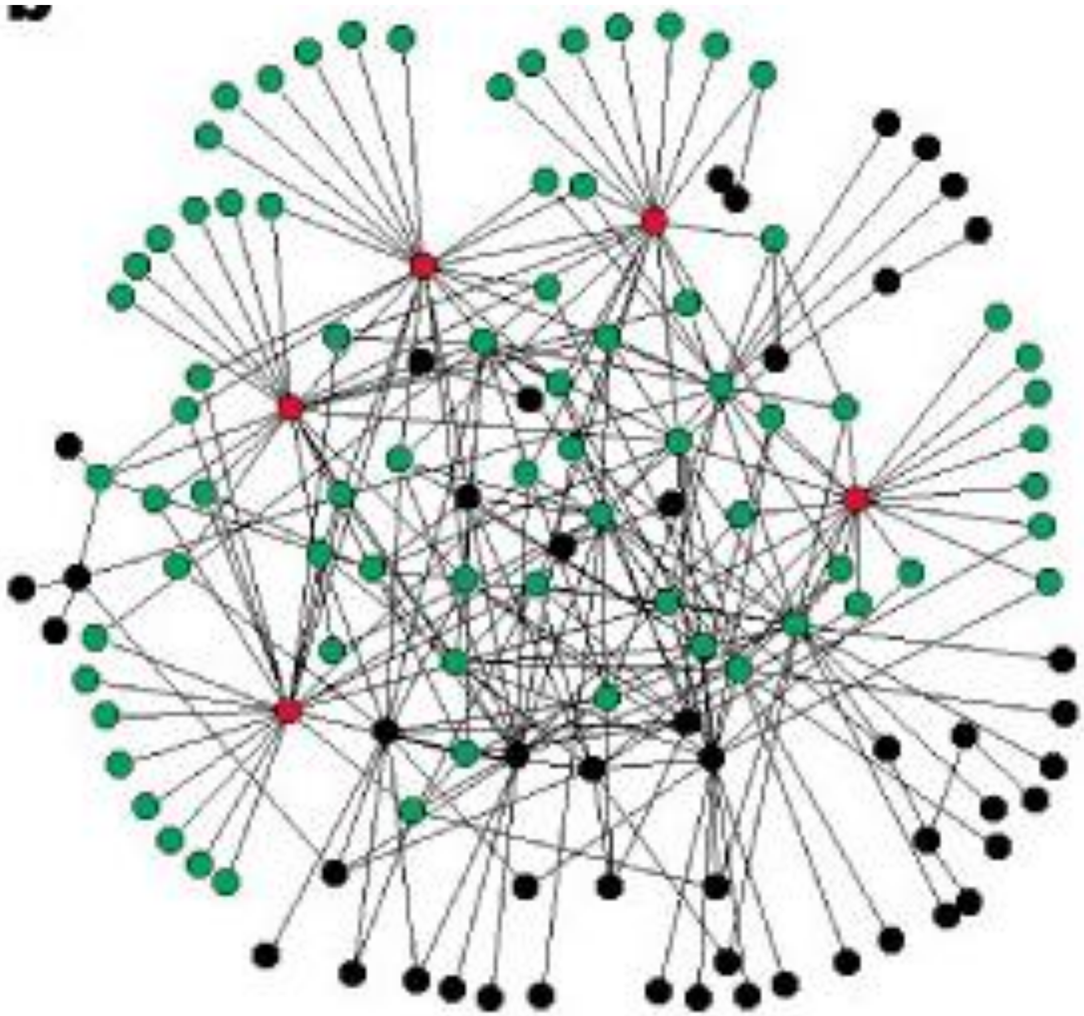
$$\sigma^2 = Var(j) = E\left[\left(j - E[j]\right)^2\right] = E\left[j^2\right] - \left(E[j]\right)^2$$

$$c = \frac{E\left[j^2\right]}{E[j]} = \frac{\left(E\left[j^2\right] - E[j]^2\right) + E[j]^2}{E[j]} = \frac{\sigma^2 + m^2}{m} = m + \frac{\sigma^2}{m}$$

$$R_0 = \beta c D = \beta \frac{E\left[j^2\right]}{E[j]} D = \beta\left(m + \frac{\sigma^2}{m}\right) D$$

R_0 rises proportional to the coefficient of variation (ratio of the variance to mean)!

Scale-Free Networks



Albert, Jeong and Barabási, Nature 406, 378-382(27 July 2000)

Scale-Free Networks

- A node's number of connections (a person's # of contacts) is denoted *k*
- The chance of having k partners is proportional to $k^{-\gamma}$.
- For human sexual networks, γ is between 2 and 3.5
  - E.g. if γ=2, likelihood having 2 partner is proportional to ¼, of having 3 is proportional to 1/9, etc.
- NB: It appears that AnyLogic's algorithm (from Barabasi & Albert *Science* 1999) imposes a γ of ~3

Power Law Scaling

- This frequency distribution is a "power law" that exhibits invariance to scale
- Suppose we "zoom in" in terms of x by a factor of α

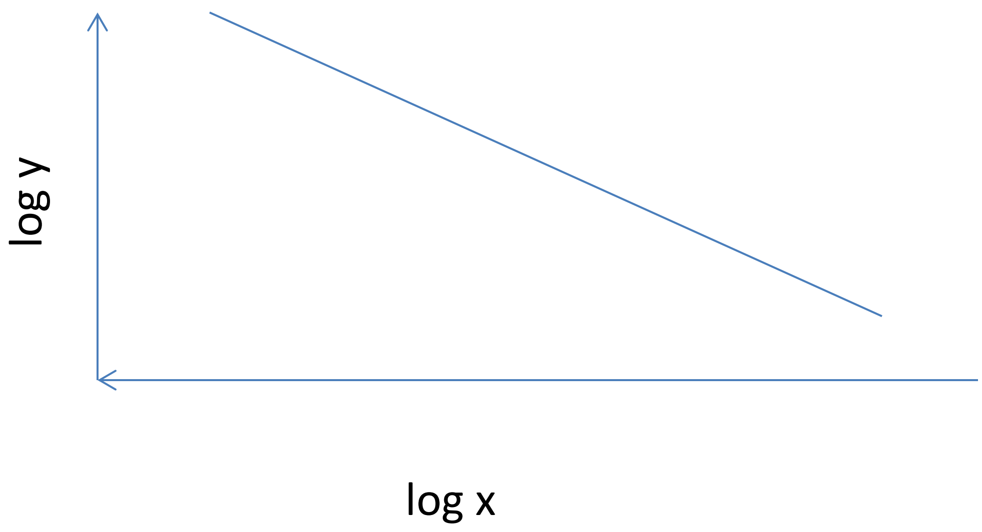
  Cf: $p(x)=cx^{-\gamma}$

  $p(\alpha x)=c(\alpha x)^{-\gamma}=c\alpha^{-\gamma}x^{-\gamma} =\alpha^{-\gamma}cx^{-\gamma} =dp(x)$

  In other words, the function p(x) "looks the same" at any scale – it is just multiplied by a different constant
- We can get power law scaling from many sources; a key source is dimensional structure
- Power law probability distributions have "long tails" compared to e.g. an exponential or normal

Recall: Power Law Scaling & Log-Log Graphs

- $y=x^a$
- $\log y = a \log x$
- If x is negative, have something like



The Signature of a Power Law

- Plotting a power law function on a log-log plot will yield a straight line
- Cf: $p(x)=cx^{-\gamma}$=>$\log p(x)=c-\gamma\log x$
- This relates to the fact that the impact of scaling (scaling) is always the identical (divides the function by the same quantity)
  - e.g. if $\gamma=2$, doubling x always divides p(x) by 4 (no matter what x is!)
  - e.g. if $\gamma=3$, doubling x always divides p(x) by 8

Power Law Scaling

- This frequency distribution is a “power law” that exhibits invariance to scale
- Suppose we change our scale (“zoom out”) in terms of number of connections (k) by a factor of α

  Cf: $p(k)=ck^{-\gamma}$

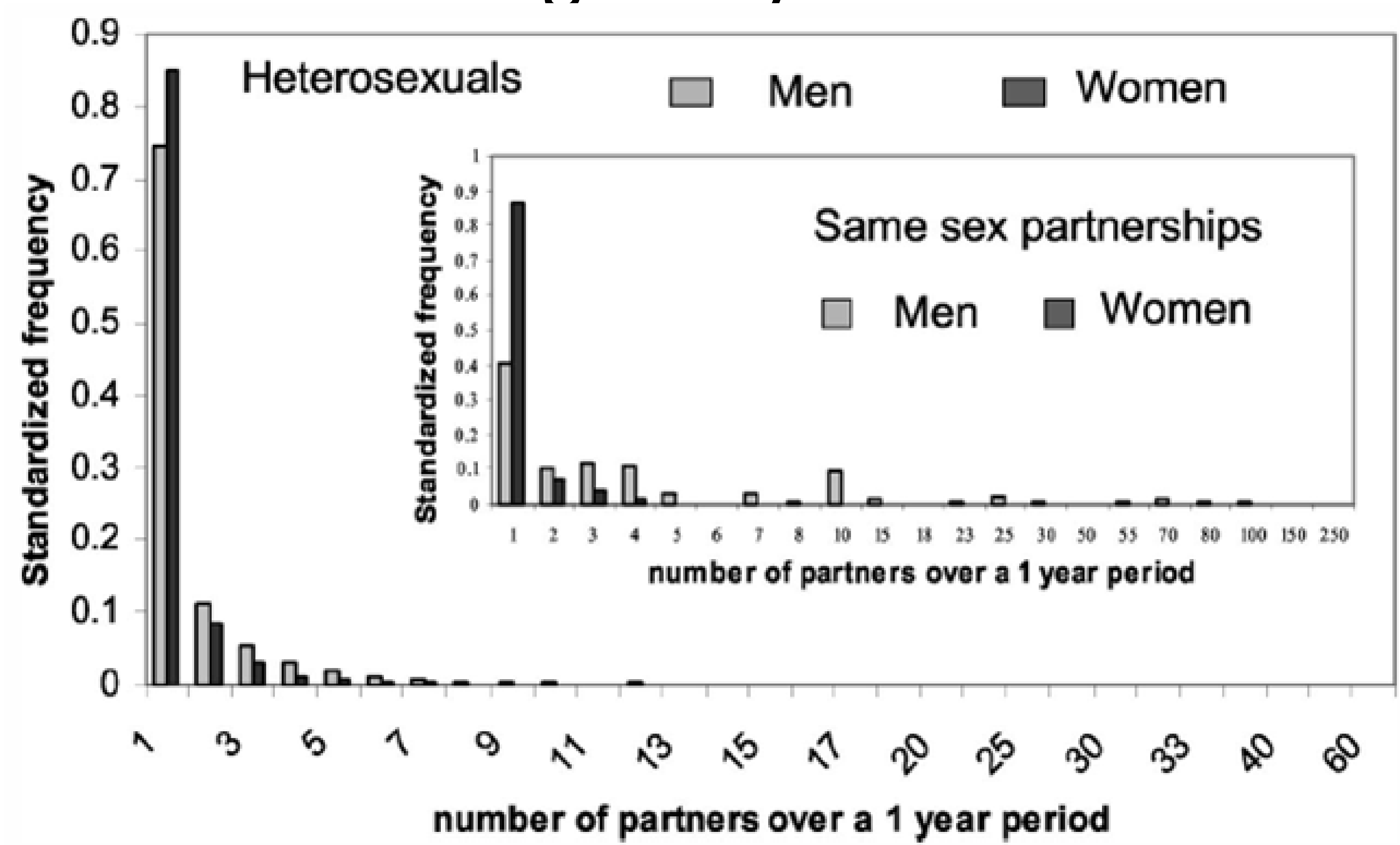
  $p(\alpha k)=c(\alpha k)^{-\gamma}=c\alpha^{-\gamma}k^{-\gamma}=\alpha^{-\gamma}ck^{-\gamma}=dp(k)$

  In other words, the function p(k) “looks the same” at any scale – it “zooming out” on the scale of # of connections by factor α just leads it to be multiplied by a different constant
- We can get power law scaling from many sources; a key source is dimensional structure
- Power law probability distributions have “long tails” compared to e.g. an exponential or normal

The Signature of a Power Law

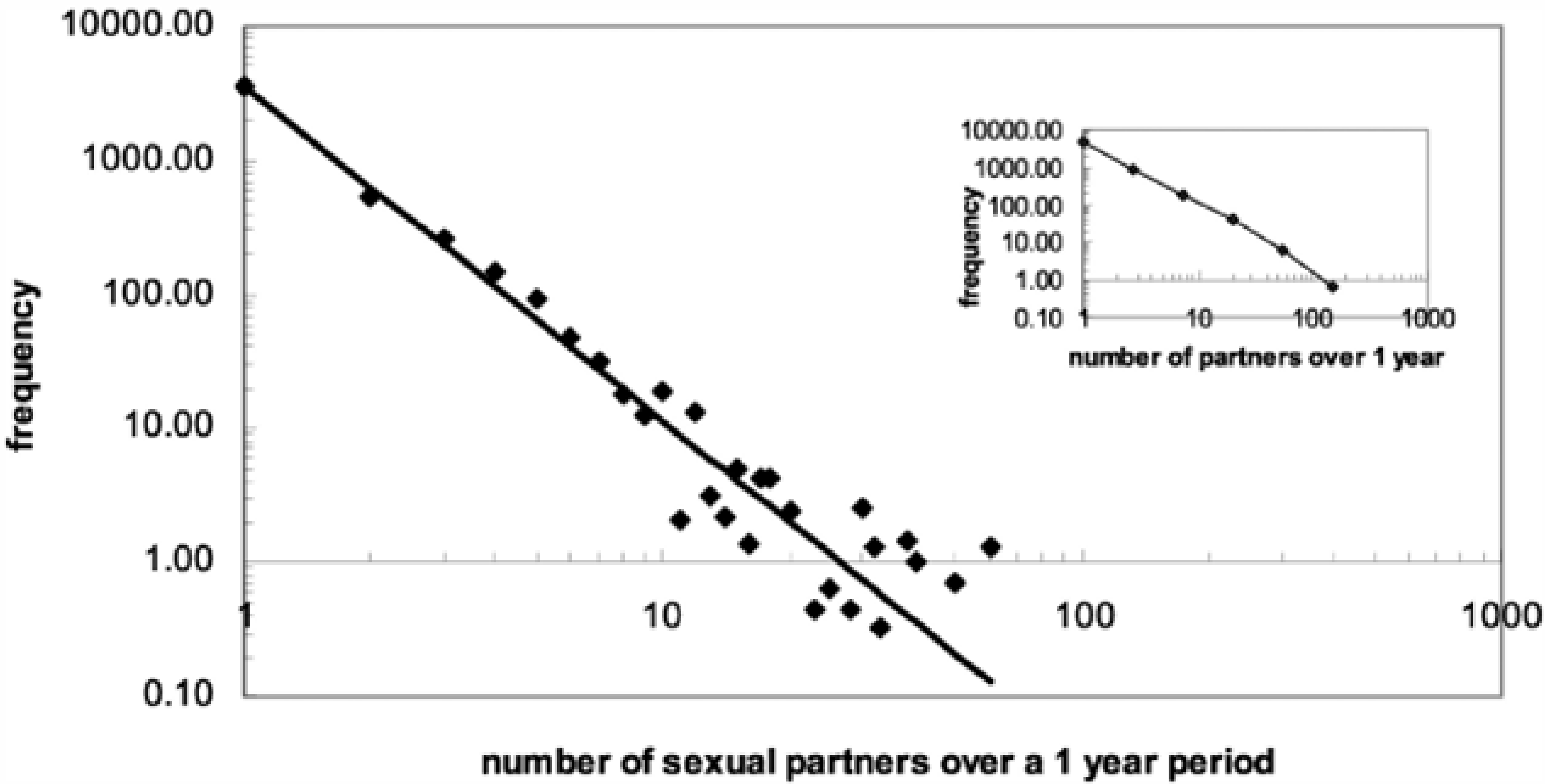
- Plotting a power law function on a log-log plot will yield a straight line
  - This reflects fact that $p(k)=ck^{-\gamma} \Rightarrow \log[p(k)]=c-\gamma\log[k]$
  - So if our axes are $v=\log[p(k)]$ and $h=\log[k]$, $v=c-\gamma h$
- This relates to the fact that the impact of scaling (scaling) is always the identical (divides the function by the same quantity)
  - e.g. if $\gamma=2$, doubling k always divides p(k) by 4 (no matter what k is!)
    - $\exists$ 4 times as many people with n connections as with 2n connections – no matter how big n is
  - e.g. if $\gamma=3$, doubling k always divides p(k) by 8

Recall: Heterogeneity in Contact Rates



Schneeberger et al., Scale-Free Networks and Sexually Transmitted Diseases:
A Description of Observed Patterns of Sexual Contacts in Britain and Zimbabwe
, Sexually Transmitted Diseases, June 2004, Volume 31, Issue 6, pp 380-387

Associated Log-Log Graph



Schneeberger et al., Scale-Free Networks and Sexually Transmitted Diseases:
A Description of Observed Patterns of Sexual Contacts in Britain and Zimbabwe
, Sexually Transmitted Diseases, June 2004, Volume 31, Issue 6, pp 380-387

Deriving the Probability Distribution Function For Scale-Free Networks

$$\sum_{k=1}^{\infty} k^{-\gamma} \approx \int_{x=1}^{\infty} x^{-\gamma} dx = \frac{1}{-\gamma+1} x^{-\gamma+1} \Big|_{1}^{\infty} = \frac{1}{-\gamma+1}(0-1) = \frac{1}{\gamma-1}$$

- PDF is $(\gamma-1)x^{-\gamma}$

Mean

- Mean

$$\int_{x=1}^{\infty} xp(x)dx = \int_{x=1}^{\infty} x(\gamma-1)x^{-\gamma}dx = \int_{x=1}^{\infty} (\gamma-1)x^{-\gamma+1}dx = \frac{\gamma-1}{\gamma-2}$$

- Variance

$$\int_{x=1}^{\infty} x^2 p(x)dx = \int_{x=1}^{\infty} x^2(\gamma-1)x^{-\gamma}dx = (\gamma-1)\int_{x=1}^{\infty} x^{-\gamma+2}dx = \frac{\gamma-1}{-\gamma+3}x^{-\gamma+3}\Big|_1^{\infty} = \frac{\gamma-1}{\gamma-3}$$

$$\sigma^2 = E[x^2] - E[x]^2 = \frac{\gamma-1}{\gamma-3} - \left(\frac{\gamma-1}{\gamma-2}\right)^2$$

Only valid if γ>3!

Variance of Human Scale-Free Networks

- For γ<3, the variance of the degree distribution for an infinitely large population is infinite! (dies off too slow)
- Recall:

$$R_0 = \beta c D = \beta \frac{E\left[j^2\right]}{E\left[j\right]} D = \beta\left(m + \frac{\sigma^2}{m}\right) D$$

- Implications
  - For a Poisson network, σ^2=m and c barely increases
  - For a scale free network with a sufficiently large population, R_0 will always be >1!
    - The disease will not die out, even if most people have low # partners!