R_0

Ottar N. Bjørnstad

Primacy of R_0

For directly transmitted pathogens, R_0 is per definition the expected number of secondary cases that arise from a typical infectious index case in a completely susceptible host population. R_0 plays a critical role for a number of aspects of disease dynamics and is therefore the focus of much study in historical and contemporary infectious disease dynamics. For perfectly immunizing infections in homogeneously mixing populations these include:

- The threshold for pathogen establishment. When R_0 is greater than one, a pathogen can invade. When it is smaller than one, the chain of transmission will stutter and break. For directly transmitted wildlife diseases there is often an associated critical host density for disease invasion. This has for example been estimated to be 1 red fox per km for rabies in Europe and 17 mice per ha for hantavirus in Montana.
- The threshold for vaccine-induced herd immunity. If a sufficient number of individuals are vaccinated, the effective reproduction number (R_E, the expected number of secondary cases in a partially immune population) will be below one, and the population will be resistant to pathogen invasion. The threshold is $p_c = 1 - 1/R_0$. Thus, measles with a R_0 of up to 20 requires around 95% vaccine cover for elimination and smallpox ($R_0 \simeq 5$) around 80%.
- In a closed epidemic the peak prevalence is $1 - (1 + \log R_0)/R_0$ and the early doubling time is $\log(2)V/\log R_0$, where V is the serial interval (the average infection-to-onwards-transmission time).
- The final epidemic size is given by R_0 according to the approximate relationship $f \simeq 1 - \exp(-R_0)$ if there are no changes to host behavior in response to the epidemic.
- In a stable host population the mean age of infection is approximately $\bar{a} \simeq L/(R_0 - 1)$, where L is host life-expectancy.
- The susceptible fraction at equilibrium is $S^* = 1/R_0$. A consequence of this is that for competing strains that elicit cross-protecting immunity R_0 will determine competitive dominance and strain replacement. A recent illustration of this is the replacement among SARS-CoV-2 variants as evolution increases human-to-human transmission.

For these reasons and more, a lot of attention has been given to estimating R_0 for various infectious diseases.

Preamble: rates and probabilities When considering events in modeling terms, a rate x per time unit is defined on $[0, \infty]$ and $1/x$ is the average time to an event (if the rate remains constant). If events are random and independent, the probability of no events in a time interval Δt is $\exp(-x\Delta t)$ and the number of events in Δt follows a Poisson distribution with mean $x\Delta t$ (if the rate remains constant). A probability, in contrast, is defined on $[0, 1]$. If we observe a probability p of something happening in a time interval, we can back-calculate the associated (constant) rate as $x = -log(1 - p)/\Delta t$.

If there are two competing processes, with rate x at which event one (eg recovery) happens and rate y at which event two (eg death) happens, the probability of ending up with outcome one is $x/(x + y)$ and the probability of ending up with outcome two is $y/(x + y)$. This scales such that with three competing rates the probability of outcome one is $x/(x + y + z)$, etc.

Estimation: Regression Consider the weekly measles data from the 2003 outbreak in Niamey, Niger . The data is available as the *niamey* dataset in the *epimdr2* package. The *tot_cases* column represents the total incidence across the city for each week of the outbreak.

```
require(epimdr2)
```

```
## Loading required package: epimdr2
```

```
## Loading required package: shiny
```

```
## Loading required package: deSolve
```

```
## Loading required package: plotly
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

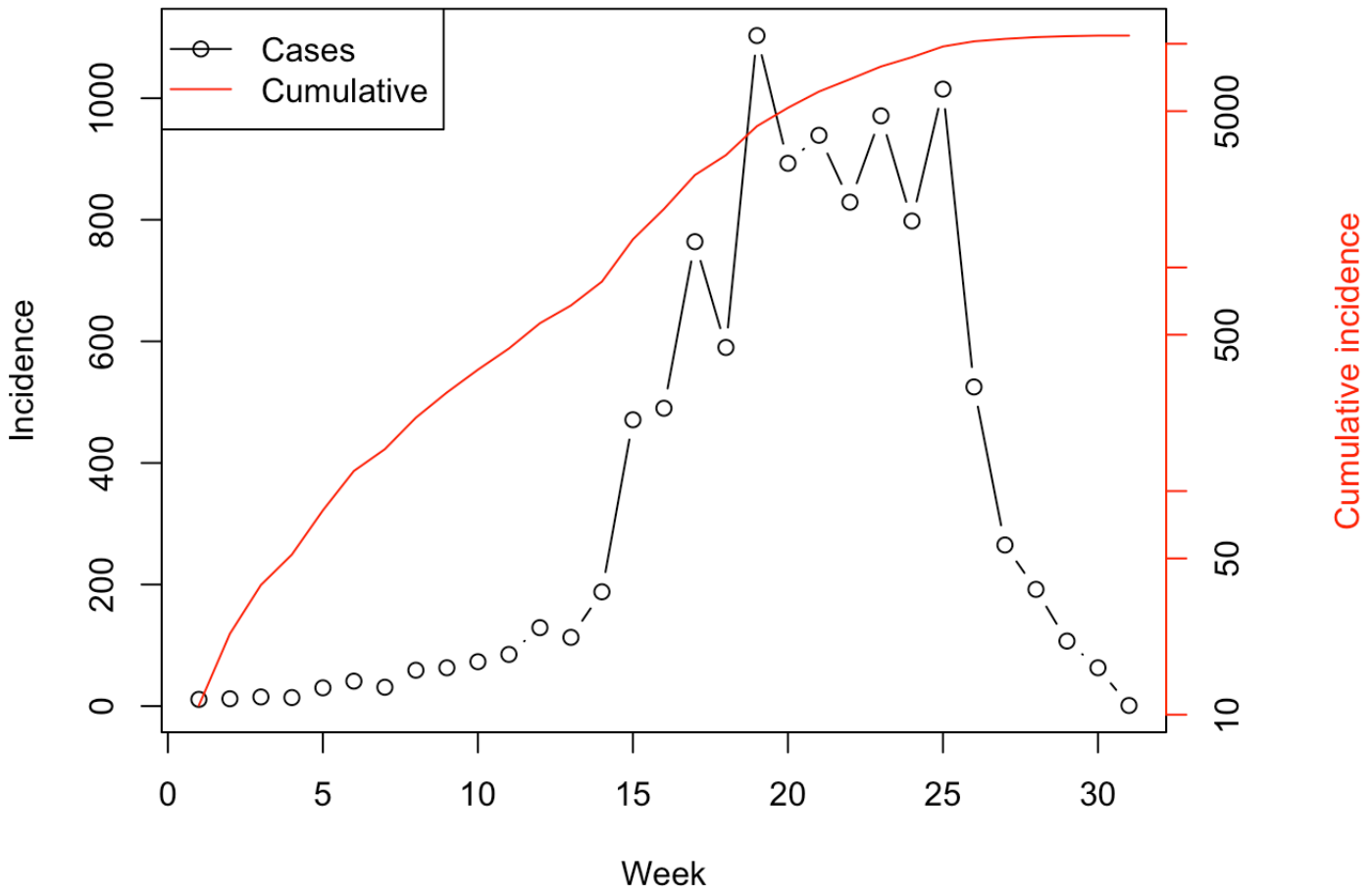
```
## Loading required package: polspline
```

```
data(niamey)
head(niamey[,1:5])
```

```
##   absweek week tot_cases tot_mort lethality
## 1       1   45        11        0  0.000000
## 2       2   46        12        1  8.333333
## 3       3   47        15        0  0.000000
## 4       4   48        14        1  7.142857
## 5       5   49        30        0  0.000000
## 6       6   50        41        1  2.439024
```

The following log-plot provides a visual inspection to identify the initial period of exponential growth.

```
par(mar = c(5,5,2,5))
plot(niamey$absweek, niamey$tot_cases, type = "b",
  xlab = "Week", ylab = "Incidence")
par(new = TRUE)
plot(niamey$absweek, niamey$cum_cases, type = "l",
  col = "red", axes = FALSE, xlab = NA,
  ylab = NA, log = "y")
axis(side = 4, col="red")
mtext(side = 4, line = 4, "Cumulative incidence", col="red")
legend("topleft", legend = c("Cases", "Cumulative"),
  lty = c(1,1), pch = c(1,NA), col = c("black", "red"))
```



The simplest idea is that during initial spread susceptible depletion may be sufficiently negligible that the epidemic may be assumed to grow in an exponential fashion. Basic ecology says that the rate of exponential growth is $r = \log(R_0)/V$, where V is the generation time; thus $R_0 = \exp(rV)$. The time for a population to double is $\log(2)/r$. For pathogens, the V represents the *serial interval* which is the average time between infection and reinfection.

The nature of exponential growth is such that incidence growth is $\exp(r)$ where $r = (R_0 - 1)/V$. The simplest way to estimate R_0 is thus to regress log(cumulative incidence) on time to estimate the rate of exponential increase (r) and then back-calculate $R_0 = Vr + 1$

The cumulative incidence looks pretty log-linear for the first 6 weeks or so. The data is weekly and the serial interval for measles is around 10–12 days, thus V is around 1.5–1.8 weeks. We can calculate R_0 assuming either 1.5 or 1.8:

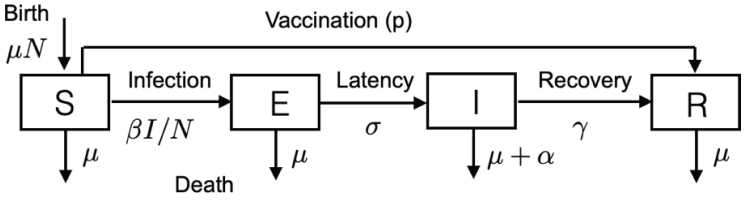
```
fit=lm(log(cum_cases)~absweek, subset=absweek<7,
   data=niamey)
r=fit$coef["absweek"]
V=c(1.5, 1.8)
V*r+1
```

```
## [1] 1.694233 1.833080
```

So a fast-and-furious estimate of the reproduction number for this outbreak places it in the 1.5–2 range. Measles exhibits recurrent epidemics in the presence of various vaccination campaigns in Niger, so this number represents an estimate of the *effective} reproduction number*, R_E, at the beginning of this epidemic.

Estimation: The logical method

For the simple SIR $R_0 = \beta/(\gamma + \mu)$. This is the correct quantity assuming that the force-of-infection (the rate at which susceptibles are infected) is $\beta I/N$, there is no latent period and no disease-induced mortality, so the index case is expected to be infectious for a period of $1/(\gamma + \mu)$ time units during which it will transmit at a rate of $\beta * N/N$. The numerator comes about because all the N individuals in the population is by definition susceptible when we consider the basic reproduction number, thus initial $S = N$.



Different SIR-like flows will produce different quantifications of R_0 but we can use the same logic for all linear flows. Consider, for example, the case when infections have a latent period leading to the SEIR model of the flow of hosts between usceptible, xposed (but not yet infectious), nfectious and emoved (either recovered with immunity or dead) compartments in a randomly mixing population:

$$\frac{dS}{dt} = \underbrace{\mu N(1-p)}_{\text{recruitment}} - \underbrace{\beta I \frac{S}{N}}_{\text{infected}} - \underbrace{\mu S}_{\text{dead}}$$

$$\frac{dE}{dt} = \underbrace{\beta I \frac{S}{N}}_{\text{infected}} - \underbrace{\sigma E}_{\text{infectious}} - \underbrace{\mu I}_{\text{dead}}$$

$$\frac{dI}{dt} = \underbrace{\sigma E}_{\text{infectious}} - \underbrace{\gamma I}_{\text{recovered}} - \underbrace{(\mu + \alpha) I}_{\text{dead}}$$

$$\frac{dR}{dt} = \underbrace{\gamma I}_{\text{recovered}} - \underbrace{\mu R}_{\text{dead}} + \underbrace{\mu N p}_{\text{vaccinated}}$$

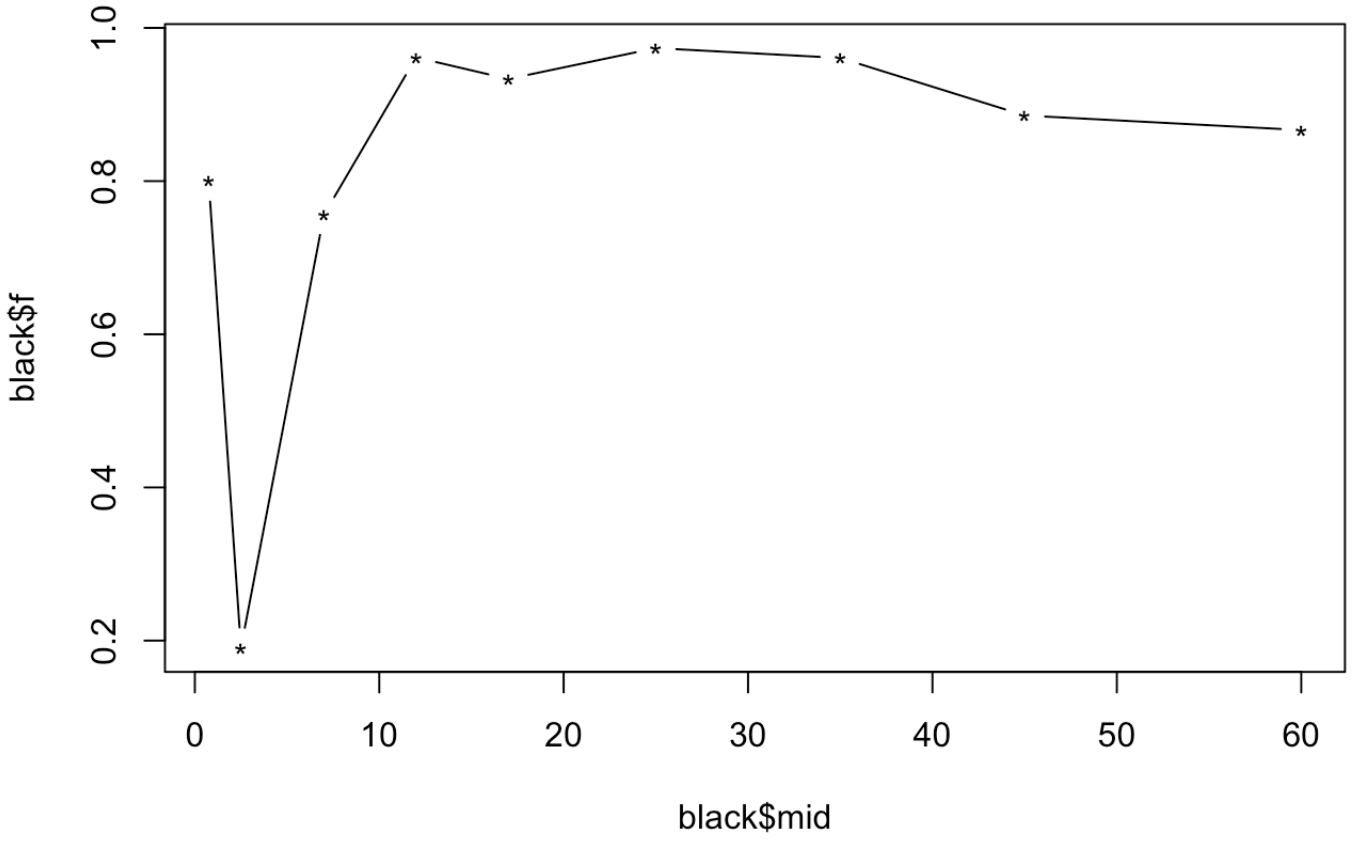
Here susceptibles are assumed either vaccinated at birth (fraction p), or infected at a rate $\beta I/N$. Infected individuals will remain in the latent class for an average period of $1/(\sigma + \mu)$ and subsequently (if they escape natural mortality at a rate μ) enter the infectious class for an average time of $1/(\gamma + \mu + \alpha)$; α is the of disease induced mortality (case fatality rate). By the rules of competing rates, the case fatality rate is $\alpha/(\gamma + \mu + \alpha)$ because during the time an individual is expected to remain in the infectious class the disease is killing them at a rate α. By a similar logic, the probability of recovering with immunity (for life in the case of the SEIR model) is $\gamma/(\gamma + \mu + \alpha)$. Putting all these pieces together and assuming no vaccination, the

expected number of secondary cases in a completely susceptible population is: probability of making it through latent stage without dying * expected infectious period * transmission rate while infectious. Thus, $R_0 = \frac{\sigma}{\sigma+\mu}\frac{1}{\gamma+\mu+\alpha}\frac{\beta N}{N} = \frac{\sigma}{\sigma+\mu}\frac{\beta}{\gamma+\mu+\alpha}$.

Estimation: Mean age of infection

For endemic fully immunizing infections in a constant-sized host population, R_0 is related to mean age of infection, $\bar{a}$, according to $R_0 \simeq 1 + L/\bar{a}$ where L is the life expectancy of the host. This rule of thumb is often used in conjunction with seroprevalence-by-age profiles to get estimates of R_0. For example Black's measles seroprevalence data:

```
plot(black$mid, black$f, pch = "*", type="b")
```



Mean age of infection is around 6

```
R0=75/6
R0
```

```
## [1] 12.5
```

So R_0 is around 12.5

Estimation: Contact Tracing

Contact tracing can provide direct estimates of R_0. showed that this together with size-distributions of subcritical transmission chains can provide estimates in important low R_0 ("subcritical") settings, such as human monkey pox in the face of eroding smallpox herd-immunity. They estimated the human-to-human reproduction number to be 0.32. Given that the smallpox vaccine is likely to be cross-protective against monkey pox, the worry is that this effective reproduction number will increase over time since smallpox vaccination is no longer carried out. Contact tracing was also used to estimate R_0 during the early spread of SARS during the 2003 outbreak . The the type of branching process models used by, for example, and will be discussed further in Sect. $???$.

De (2004) did a contact-tracing study of the spread of gonorrhea across a sexual network in Alberta, Canada. The directional transmission graph among the 89 individuals is in the *gonnet*}* dataset. The initial cluster of 17 cases all frequented the same bar, each infected between 0 and 7 other partners with 2.17 as the average. The subsequent infections, in turn, infected between 0 and 6 partners with an average of 0.62. The drop is (i) due to the sexual network being depleted of susceptibles, and (ii) because infection across heterogenous networks will differentially infect individuals according to their number of contacts. The statnet package has great tools for visualizing chains of transmission on networks.

```
require(statnet)
```

```
## Loading required package: statnet
```

```
## Loading required package: tergm
```

```
## Loading required package: ergm
```

```
## Loading required package: network
```

```
##
## 'network' 1.18.1 (2023-01-24), part of the Statnet Project
## * 'news(package="network")' for changes since last version
## * 'citation("network")' for citation information
## * 'https://statnet.org' for help, support, and other information
```

```
##
## 'ergm' 4.5.0 (2023-05-27), part of the Statnet Project
## * 'news(package="ergm")' for changes since last version
## * 'citation("ergm")' for citation information
## * 'https://statnet.org' for help, support, and other information
```

```
## 'ergm' 4 is a major update that introduces some backwards-incompatible
## changes. Please type 'news(package="ergm")' for a list of major
## changes.
```

```
## Loading required package: networkDynamic
```

```
##
## 'networkDynamic' 0.11.3 (2023-02-15), part of the Statnet Project
## * 'news(package="networkDynamic")' for changes since last version
## * 'citation("networkDynamic")' for citation information
## * 'https://statnet.org' for help, support, and other information
```

```
## Registered S3 method overwritten by 'tergm':
##   method                   from
##   simulate_formula.network ergm
```

```
##
## 'tergm' 4.2.0 (2023-05-30), part of the Statnet Project
## * 'news(package="tergm")' for changes since last version
## * 'citation("tergm")' for citation information
## * 'https://statnet.org' for help, support, and other information
```

```
##
## Attaching package: 'tergm'
```

```
## The following object is masked from 'package:ergm':
##
##     snctrl
```

```
## Loading required package: ergm.count
```

```
##
## 'ergm.count' 4.1.1 (2022-05-24), part of the Statnet Project
## * 'news(package="ergm.count")' for changes since last version
## * 'citation("ergm.count")' for citation information
## * 'https://statnet.org' for help, support, and other information
```

```
## Loading required package: sna
```

```
## Loading required package: statnet.common
```

```
##
## Attaching package: 'statnet.common'
```

```
## The following object is masked from 'package:ergm':
##
##     snctrl
```

```
## The following objects are masked from 'package:base':
##
##     attr, order
```

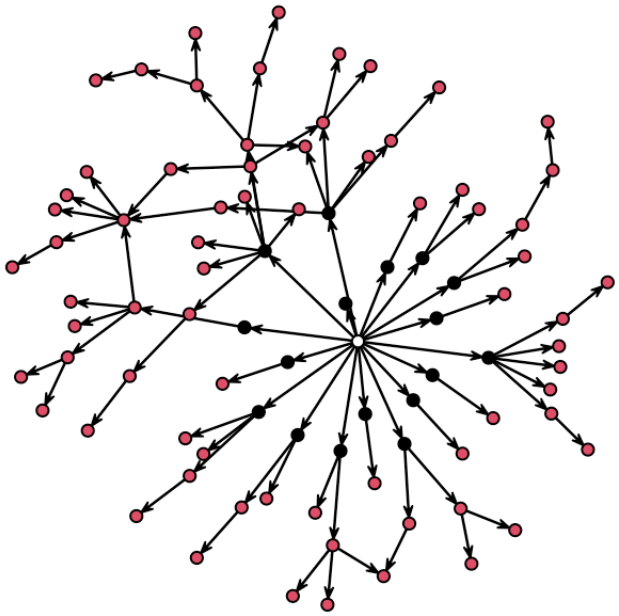
```
## sna: Tools for Social Network Analysis
## Version 2.7-1 created on 2023-01-24.
## copyright (c) 2005, Carter T. Butts, University of California-Irvine
##  For citation information, type citation("sna").
##  Type help(package="sna") to get started.
```

```
## Loading required package: tsna
```

```
##
## 'statnet' 2019.6 (2019-06-13), part of the Statnet Project
## * 'news(package="statnet")' for changes since last version
## * 'citation("statnet")' for citation information
## * 'https://statnet.org' for help, support, and other information
```

```
## unable to reach CRAN
```

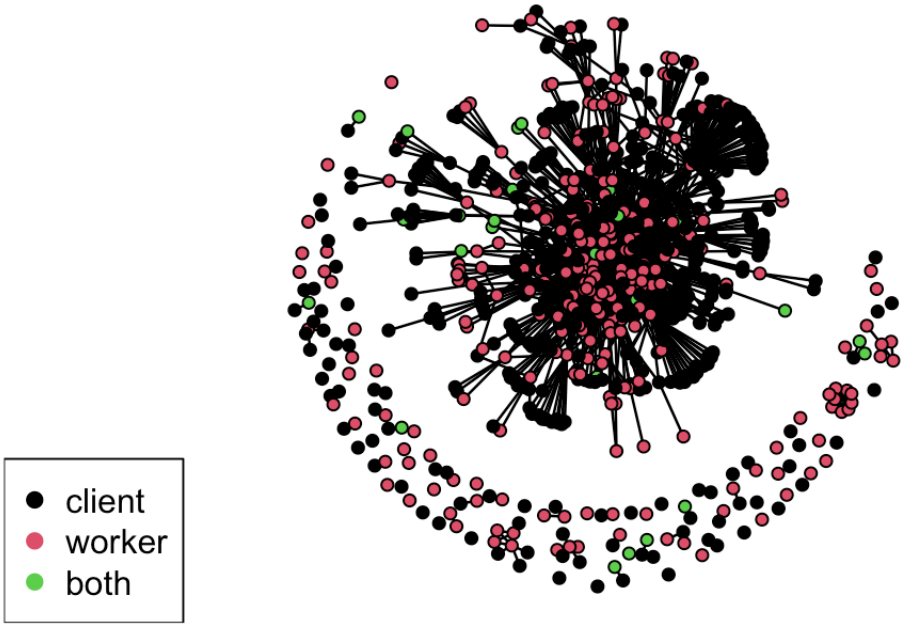
```
data(gonnet)
nwt=network(gonnet, directed=TRUE)
plot(nwt, vertex.col=c(0, rep(1,17), rep(2,71)))
```



White is bar. Black are ``first" generation cases and read are onwards.

Post script: Social heterogeneities: The P90 project mapped sexual contacts among $5,493$ individuals in Colorado springs between 1988 and 1990. The study was motivated by the need to assess spread of STDs from risky sexual activities in the face of the then rising HIV pandemic. The *csprin* data on participation role in the network (*role: 1=client, 2=worker, 3=both*) and the 749-by-749 binary matrix (*cm*) is mapping of contacts among the subset are included in the data list.

```
require(statnet)
data(cspring)
#convert contact matrix to network object
csnwrk = network(cspring$cm, matrix.type = "adjacency",
   directed = FALSE)
#set individual attributes to network
set.vertex.attribute(csnwrk, "role", cspring$nodes$type)
network.vertex.names(csnwrk) = c("client", "worker", "both")
plot(csnwrk, vertex.col = cspring$nodes$type)
legend("bottomleft", c("client", "worker", "both"),
   col = 1:3, pch = 21, pt.bg = 1:3)
```



A depiction of the Colorado springs sex worker and client network}

A key feature of a social network is the individual level heterogeneity in number of contacts. The violin plot is a very useful visualization of distributional heterogeneities. The figure shows the number of links of clients and workers. The workers have on average nearly 10 clients and clients just over two sex partners. Conspicuously, the distribution is heavily skewed. A small number of people has a disproportionate number of contacts as shown in the log-log plot:

```
require(vioplot)
```

```
## Loading required package: vioplot
```

```
## Loading required package: sm
```

```
## Package 'sm', version 2.2-5.7: type help(sm) for summary information
```

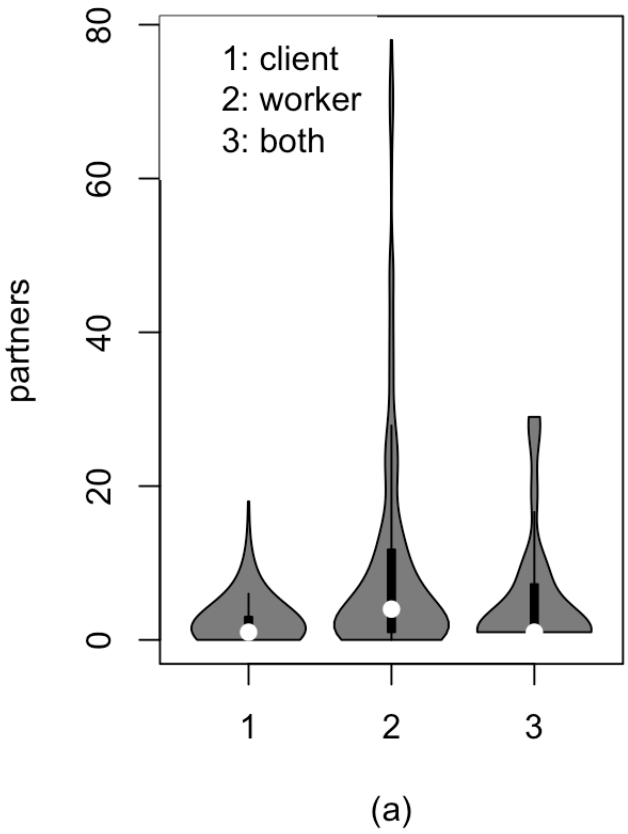
```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

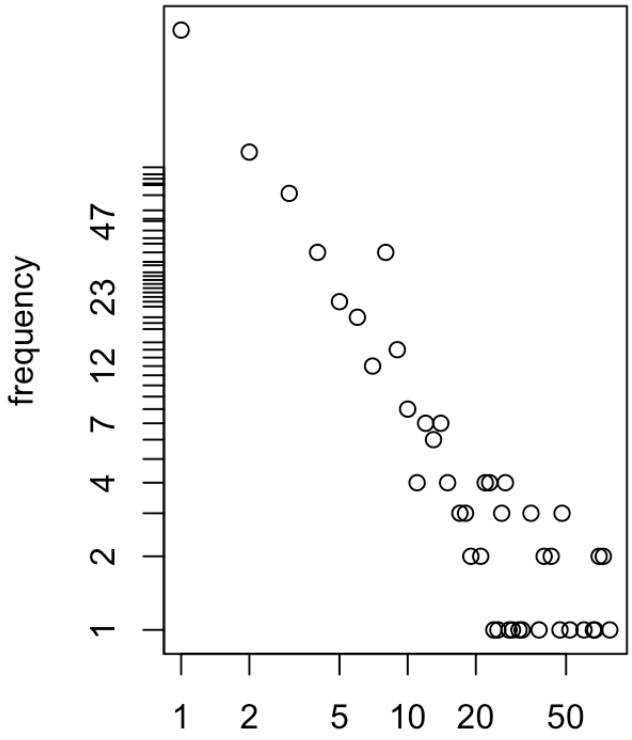
```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
par(mfrow = c(1, 2))
#violin plot
vioplot(apply(cspring$cm, 2, sum) ~ cspring$nodes$type,
   ylab = "partners", xlab = "(a)", h = 3)
legend("topleft", c("1: client", "2: worker", "3: both"), box.lty = 0)
#log-log plot
dd = table(apply(cspring$cm, 2, sum))
plot(as.numeric(names(dd)), dd, log = "xy", ylab = "frequency", xlab = "(b)")
```

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): 1 x value <= 0 omitted from
## logarithmic plot
```


partners
0
20
40
60
80
1: client
2: worker
3: both
1
2
3

(a)


frequency
1
2
4
7
12
23
47
1
2
5
10
20
50

(b)