Trajectories of Ebola infection probability in West Africa

John M. Drake

September 2, 2016

Description

During the 2013-2015 West Africa epidemic, the probability that a particular district would report its first case on a given day varied as a result of the path of the epidemic through the affected countries and the connections of each district with other districts. This shows that there exists a *network of infection risk* that develops during epidemics such that the risk faced by individual units depends not only on unit-level features, but also the features and status of unit neighbors and neighbors-of-neighbors.

Kramer et al. (2016b) developed a stochastic model for the spatial spread of Ebola in West Africa and estimated the coefficients for the transmission network that existed among 290 administrative districts. Results of that work highlighted the importance of population density, location, border closures, and long distance dispersal. A key figure from that work, reproduced here, provides a graphical depiction of both the static network of transmission potential and geographic path taken by the epidemic. What this figure does not show is the time-varying risk experienced by each administrative unit as the epidemic proceeded.

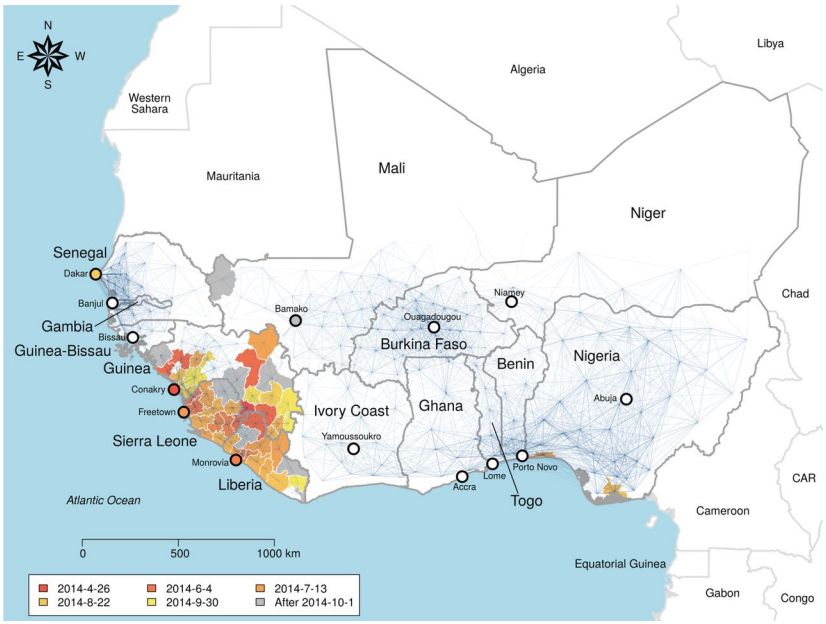

Figure 1: Potential transmission network for spread of Ebola in West Africa (Kramer et al 2016a).

Here, we evaluate the model of Kramer et al. (2016b) to produce estimates of the time-varying daily risk of infection. The resulting trajectories of Ebola infection probability reflect both the fit model developed by Kramer et al. (2016b) and the actual change of infection status through time of all districts in the network. The following R code depends on the packages `maptools`, `PBSmapping`, and `Matrix` and the R script `ebola-GravNet-border.R` from the Dryad Repository Kramer et al. (2016a) (available online: http://dx.doi.org/10.5061/dryad.k95j3). This code produces two files. First, `data.csv` is a comma separated file containing county names, an individual identifier, a region identifier, geographic coordinates, population size, and day number for infection acquisition and recovery for up to two episodes of active transmission for all districts. Second, `table.csv` contains the daily infection probability for the 290 district from April 26, 2014 to November 10, 2014.

Code

First we write some required functions to: (1) collect tabular spatial data from a shapefile, and (2) construct the table of infection probabilities over time.

```
collect.wa <- function(cutoff=TRUE){
  require(maptools)
  require(PBSmapping)

  polys <- readRDS("WestAfricaCountyPolygons.rds")
  centroids <- calcCentroid(SpatialPolygons2PolySet(polys), rollup=1)
  region <- as.integer(as.factor(polys$ISO))

  outbreak <- readRDS("OutbreakDateByCounty_Summer_AllCountries.rds")

  first <- as.integer(outbreak$first)
  first[which(is.na(first))] <- Inf
  first <- first - min(first)
  first.recover <- as.integer(outbreak$first.recover)
  first.recover[which(is.na(first.recover))] <- Inf
  first.recover <- first.recover - min(first.recover)
  second <- as.integer(outbreak$second)
  second[which(is.na(second))] <- Inf
  second <- second - min(second)
  if(cutoff){
    second[which(second>157)] <- Inf
  }
  second.recover <- as.integer(outbreak$second.recover)
  second.recover[which(is.na(second.recover))] <- Inf
  second.recover <- second.recover - min(second.recover)

  out <- data.frame(    county_names = outbreak$county_names,
            PID = centroids$PID,
            region = region,
            x = centroids$X,
            y = centroids$Y,
            pop = polys$pop.size,
            first = first,
            first.recover=first.recover,
            second=second,
            second.recover=second.recover)

  return(out)
}

overtime.table <- function(data, dist, mass, cross, b){
  tmax <- max(data[which(data!=Inf)]) + 1
  ret <- matrix(NA, nrow=length(data), ncol=tmax)
  for(i in 1:tmax){
    ret[,i] <- getprobs(data<i, dist, mass, cross, b)
  }
  return(ret)
}
```

The model is contained in the R script `ebola-GravNet-border.R` which is loaded next.

```
source('ebola-GravNet-border.R')
```

Next, we construct the tabular data and write to the file `data.csv`.

```
data <- collect.wa()
write.csv(data, file='data.csv')
```

Infection dates are contained in the R data file `OutbreakDateByCounty_Summer_AllCountries.rds`.

```
latebreak <- readRDS("OutbreakDateByCounty_Summer_AllCountries.rds")
```

Some values are needed to parameterize the model.

```
dist <- makedist(rbind(data$x,data$y))
cross <- makecross.group(data$region)
b <- c(5.7924, 105.67, 0.18597, 0.15021)
```

Next, we format the infection dates.

```
second <- as.integer(latebreak$second)
second[which(is.na(second))] <- Inf
second <- second - min(second)
```

Finally, we calculate the table of infection probabilities and store the result in the file `table.csv`.

```
table <- overtime.table(second, dist, data$pop, cross, b)
write.csv(table, file='table.csv')
```

Acknowledgements

Tomlin Pulliam contributed code. Drew Kramer provided comments on the development of these data.

References

Kramer, Andrew M., J. Tomlin Pulliam, Laura W. Alexander, Andrew W. Park, Pejman Rohani, and John M. Drake. 2016a. "Data from: Spatial spread of the West Africa Ebola epidemic." *Dryad Digital Repository*. doi:10.5061/dryad.k95j3.

———. 2016b. "Spatial spread of the West Africa Ebola epidemic." *Royal Society Open Science* 3 (8). doi:10.1098/rsos.160294.