



Photo: Vladimir Batagelj, *Pajčevina*

Multivariate Networks

Vladimir Batagelj

University of Ljubljana
FMF, Mathematics

Seminar IBMI, Ljubljana, 22. October 2004

Wednesday seminar, Ljubljana, 3. November 2004

Outline

1	Multivariate networks	1
4	Example: Fisher's irises	4
10	Multivariate networks and complexity	10
11	Intermediate solution	11
13	Example: Slovenian communes	13

Multivariate networks

Let \mathbf{U} be a *set of multivariate units* and d a *dissimilarity* on it. They determine two types of networks:

The *k nearest neighbors graph* $\mathbf{G}_N(k) = (\mathbf{U}, A)$

$(X, Y) \in A \Leftrightarrow Y$ is among the k closest neighbors of X

By setting for $a(X, Y) \in A$ its value to $w(a) = d(X, Y)$ we obtain a network.

In the case of equidistant pairs of units we have to decide – or to include them all in the graph, or specify an additional selection rule.

A special case of the k nearest neighbors graph is the *nearest neighbor graph* $\mathbf{G}_N(1)$. We shall denote by \mathbf{G}_{NN}^* the graph with included all equidistant pairs, and by \mathbf{G}_{NN} a graph where a single nearest neighbor is always selected.

The *fixed-radius neighbors graph* $\mathbf{G}_B(r) = (\mathbf{U}, E)$

$(X : Y) \in E \Leftrightarrow d(X, Y) \leq r, \quad w(X, Y) = w(Y, X) = d(X, Y)$

k nearest neighbors in R

```
k.neighbor2Net <-  
# stores network of first k neighbors for  
# dissimilarity matrix d to file fnet in Pajek format.  
function(fnet,d,k){  
  net <- file(fnet,"w")  
  n <- nrow(d); rn <- rownames(d)  
  cat("*vertices",n,"\n",file=net)  
  for (i in 1:n) cat(i," \"",rn[i], "\"\n",sep="",file=net)  
  cat("*arcs\n",file=net)  
  for (i in 1:n) for (j in order(d[i,])[1:k+1]) {  
    cat(i,j,d[i,j],"\n",file=net)  
  }  
  close(net)  
}
```

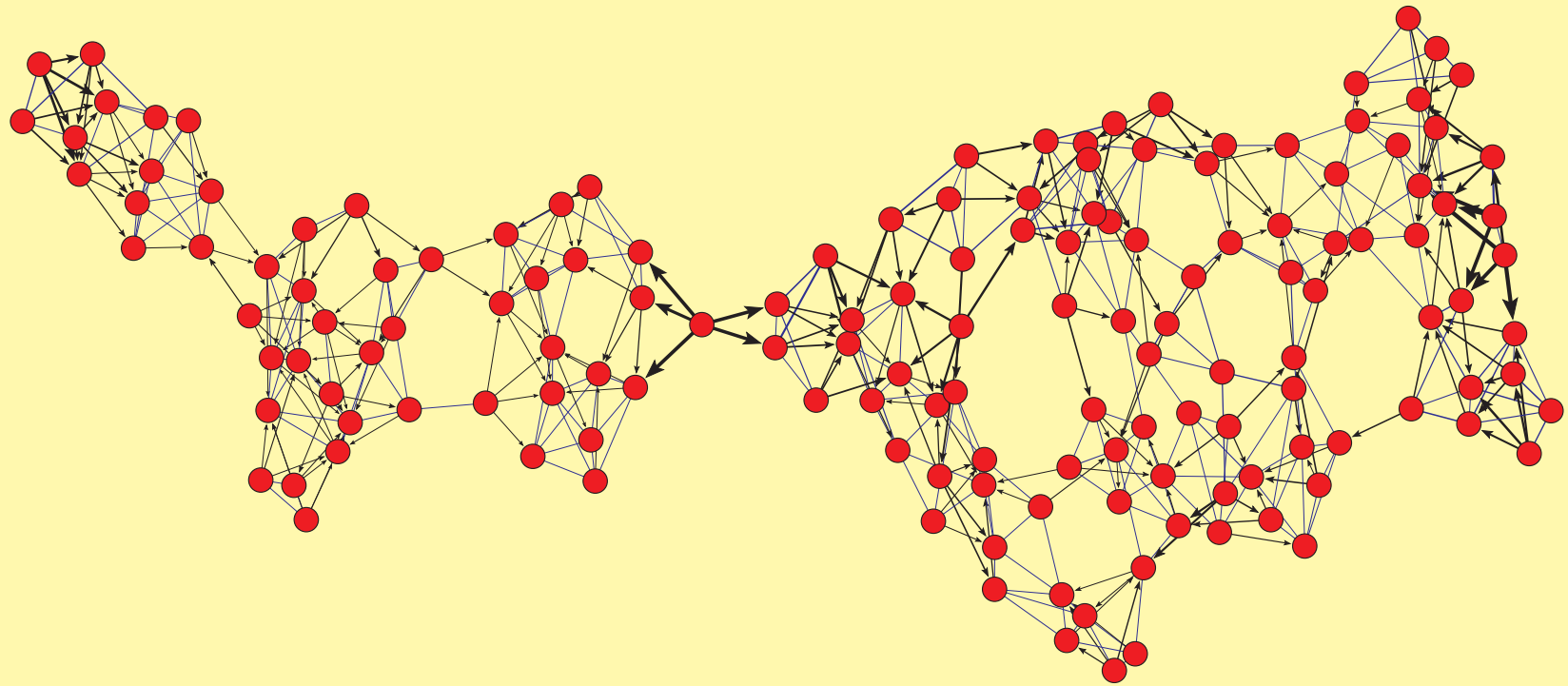
r-neighbors in R

```
r.neighbor2Net <-  
# stores network of r-neighbors (d(v,u) <= r) for  
# dissimilarity matrix d to file fnet in Pajek format.  
function(fnet,d,r){  
  net <- file(fnet,"w")  
  n <- nrow(d); rn <- rownames(d)  
  cat("*vertices",n,"\n",file=net)  
  for (i in 1:n) cat(i," \"",rn[i],"\"\\n",sep="",file=net)  
  cat("*edges\\n",file=net)  
  for (i in 1:n){  
    s <- order(d[i,]); j <- 1  
    while (d[i,s[j]] <= r) {  
      k <- s[j]; if (i < k) cat(i,k,d[i,k],"\n",file=net)  
      j <- j+1  
    }  
  }  
  close(net)  
}
```

Example: Fisher's irises

```
stand <-  
# standardizes vector x .  
function(x){  
  s <- sd(x)  
  if (s > 0) (x - mean(x))/s else x - x  
}  
  
data(iris)  
ir <- cbind(stand(iris[,1]),stand(iris[,2]),stand(iris[,3]),  
  stand(iris[,4]))  
k.neighbor2Net("iris5.net",as.matrix(dist(ir)),5)
```

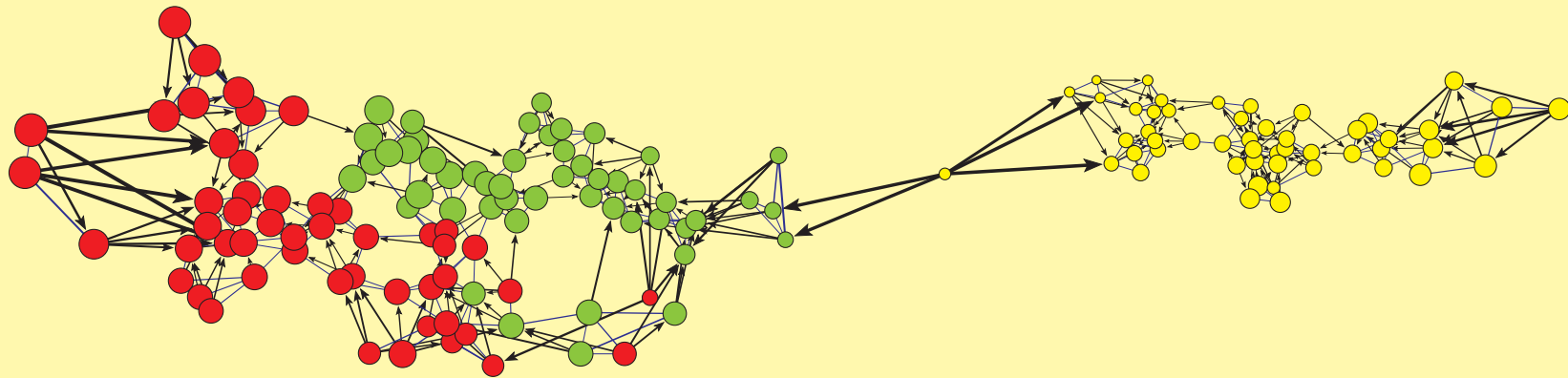
Fisher's irises / 5 nearest neighbors network



Additional network data

```
vector2Clu <-  
# stores integer vector v as Pajek partition to file fclu .  
function(fclu,v){  
  clu <- file(fclu,"w")  
  n <- length(v)  
  cat("*vertices",n,"\n",file=clu)  
  for (i in 1:n) cat(v[i],"\n",file=clu)  
  close(clu)  
}  
  
vector2Vec <-  
# stores vector v as Pajek vector to file fvec .  
function(fvec,v){  
  vec <- file(fvec,"w")  
  n <- length(v)  
  cat("*vertices ",n,"\n",file=vec)  
  for (i in 1:n) cat(v[i],"\n",file=vec)  
  close(vec)  
}  
  
vector2Clu("iris.clu",as.numeric(iris[,5]))  
vector2Vec("pl.vec",iris[,1])
```


Fisher's irises with additional data



The weights are considered in **Pajek** as dissimilarities

Draw/Options/Values of Lines/Dissimilarities

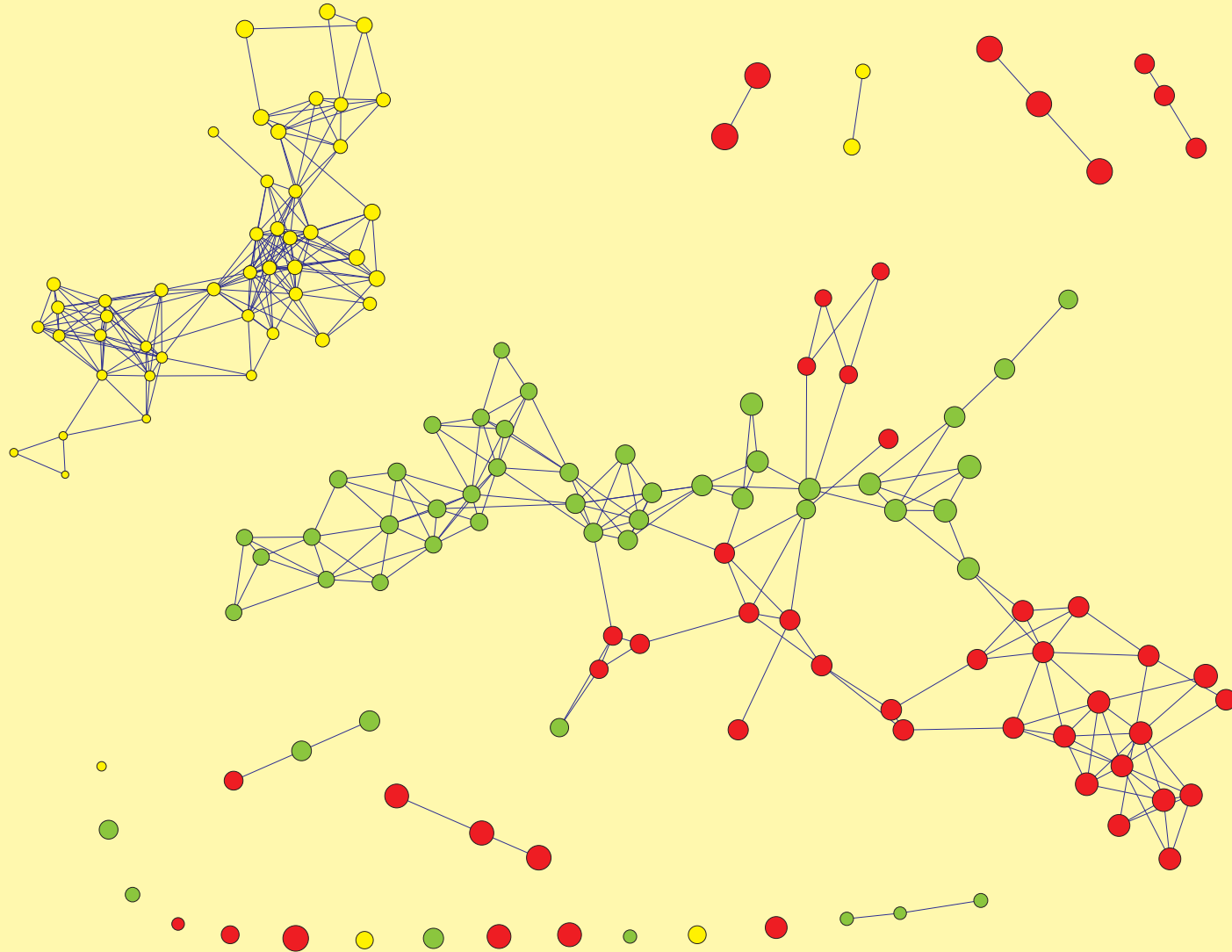
The size of vertices is proportional to $\text{iris}_1 - 4$. The color of vertices represents the clustering of irises as given in the data (iris_5).

Fisher's irises / Neighborhoods

```
r.neighbor2Net("irisR.net",d,0.5)
```

It seems that the r -neighbors network makes sense only when the 'density' of groups doesn't change.

Fisher's irises / r -neighbors network



Multivariate networks and complexity

There is no known subquadratic algorithm for determining multivariate networks. The basic paper on the subject is:

Fukunaga, K., Narendra, P.M. (1975), A branch and bound algorithm for computing k-nearest neighbors. *IEEE Transactions on Computers*, **C-24**, 750-753.

Skiena, Yianilos, Brin, Chua, Chávez &, Murtagh, Dickerson Eppstein, Kleinberg, lectures 22-24, D'haes &, CV Bib: NNC.

My *Konstanz* lectures [1, 2, 3, 4]. See [Lecture 3](#), slides 5-.

Intermediate solution

```
dist2net <-  
# Pajek network (in *arcs/*edges format) on file fnet  
# transforms into network with dissimilarities dist between  
# line endpoints as weights and store it to file fdis .  
function(fnet, fdis, dat, dist, ...){  
  net <- file(fnet,"r"); dis <- file(fdis,"w"); copy <- TRUE  
  while (copy) {  
    t <- readLines(net,n=1)  
    cat(t,"\n",file=dis)  
    if (length(t)>0){  
      if (substr(t,1,1) == "*") {  
        c <- tolower(substr(t,2,2))  
        copy <- (c != "e") & (c != "a")  
        if (copy & (c != "v")) cat("Unexpected command in NET\n")  
      }  
    } else copy <- FALSE  
  }  
}
```

... Intermediate solution

```

copy <- TRUE
digits <- c("0","1","2","3","4","5","6","7","8","9","-")
while (copy) {
  t <- readLines(net,n=1)
  if (length(t)>0){
    if (substr(t,1,1) == "*") {
      c <- tolower(substr(t,2,2))
      copy <- (c == "e") | (c == "a")
      if (copy) cat(t,"\n",file=dis)
    } else {
      z <- unlist(strsplit(t," ")); z <- z[z != ""]
      u <- as.numeric(z[1]); v <- as.numeric(z[2])
      d <- dist(rbind(dat[u,],dat[v,]),...)
      j <- 3; if (is.element(substr(z[3],1,1),digits)) j <- 4
      cat(z[1],z[2],d,file=dis)
      if (j <= length(z)) for(i in j:length(z)) cat(z[i],file=dis)
      cat("\n",file=dis)
    }
  } else copy <- FALSE
}
close(dis); close(net)
}

```

Example: Slovenian communes

Example: Slovenian communes 2002

- V_1 – name
- V_2 – average size gospodinjstva
- V_3 – average number of flats per building
- V_4 – average (netto) income

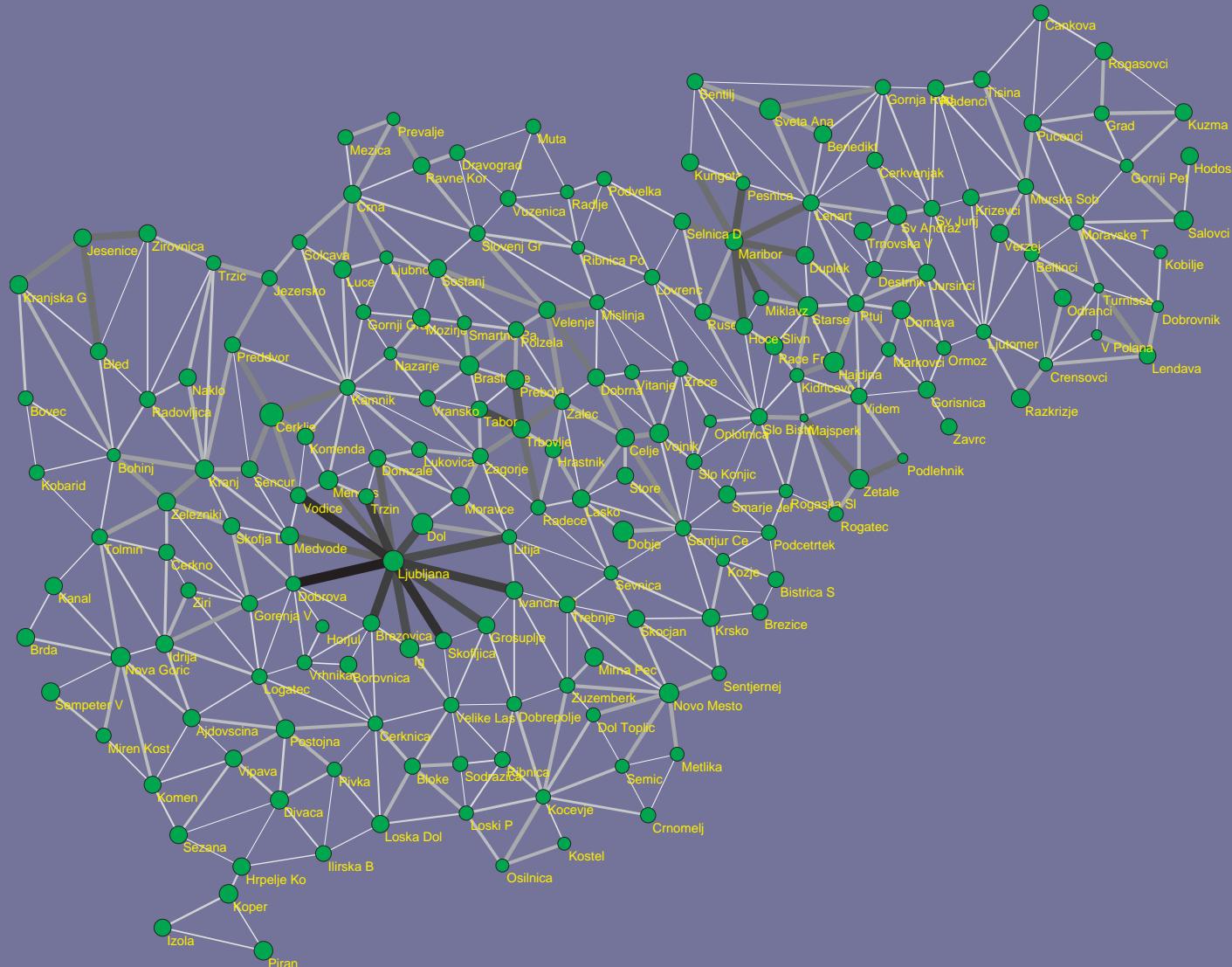
Slovenian communes – network

```
o <- read.delim("obctest.txt",header=FALSE,sep="\t",dec=",")
or <- cbind(stand(o[,2]),stand(o[,3]),stand(o[,4]))
dist2net("obcxy.net", "ObcineDis.net",or,dist)
vector2Vec("placa.vec",o[,4])

dist2net("obcxy.net", "ObcineMan.net",or,dist,method="manhattan")
```

The size of vertices is proportional to $V_4 - 70000$.

Slovenian communes



Some links

- **R-project**, **R for Windows**.
- **Pajek** – program for analysis and visualization of large networks.
- **NetAna** – programs in R and data for *Multivariate networks* (ZIP).