UNIVERSITY OF
EASTERN FINLAND

XNN Graph

Pasi Fränti, Radu Mariescu-Istodor and Caiming Zhong

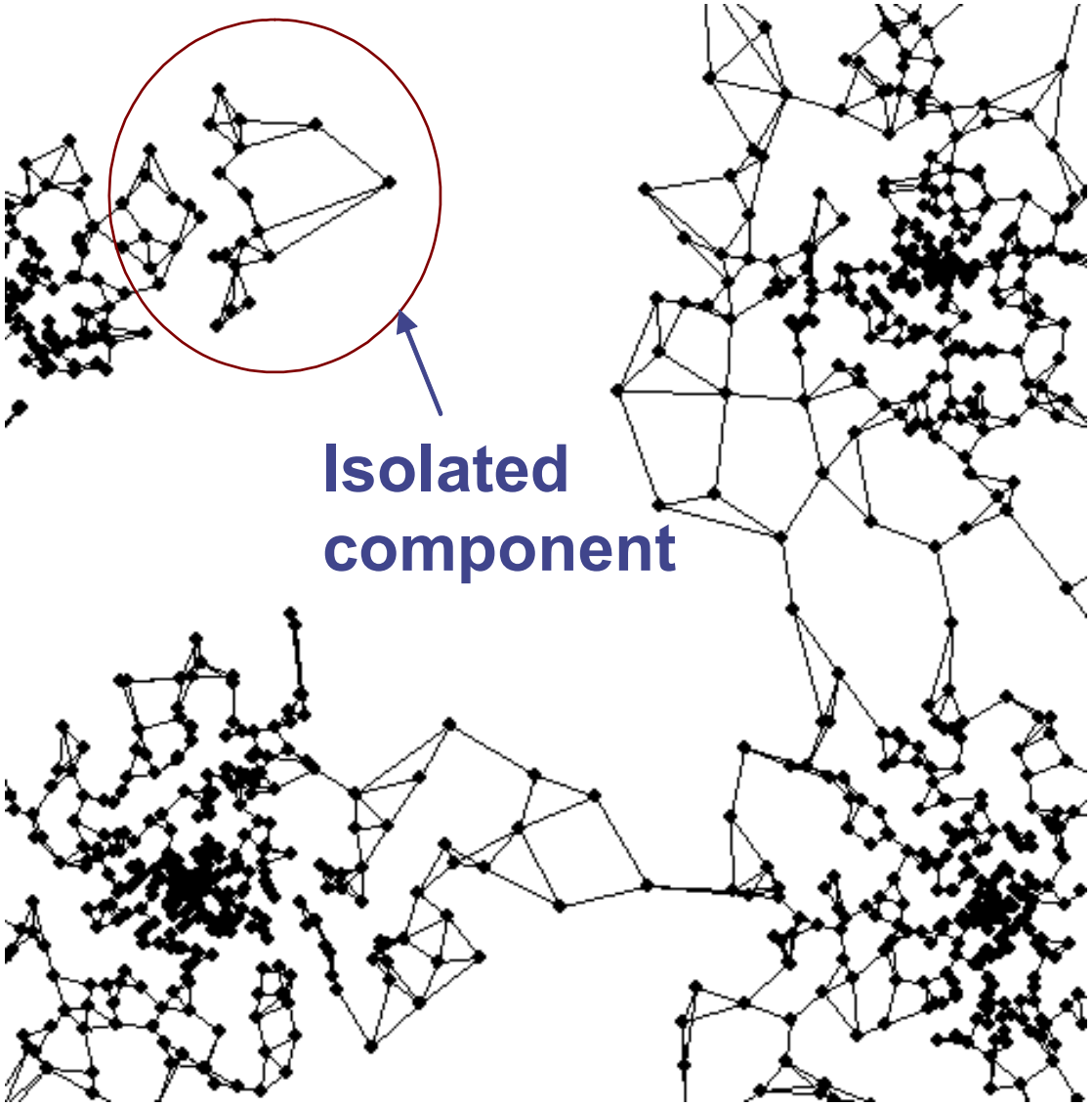
30.11.2016

Applications

- KNN classifier
- Manifold learning
- 3D object matching
- Clustering
- Outlier detection
- Traveling salesman problem
- Word similarity in web mining

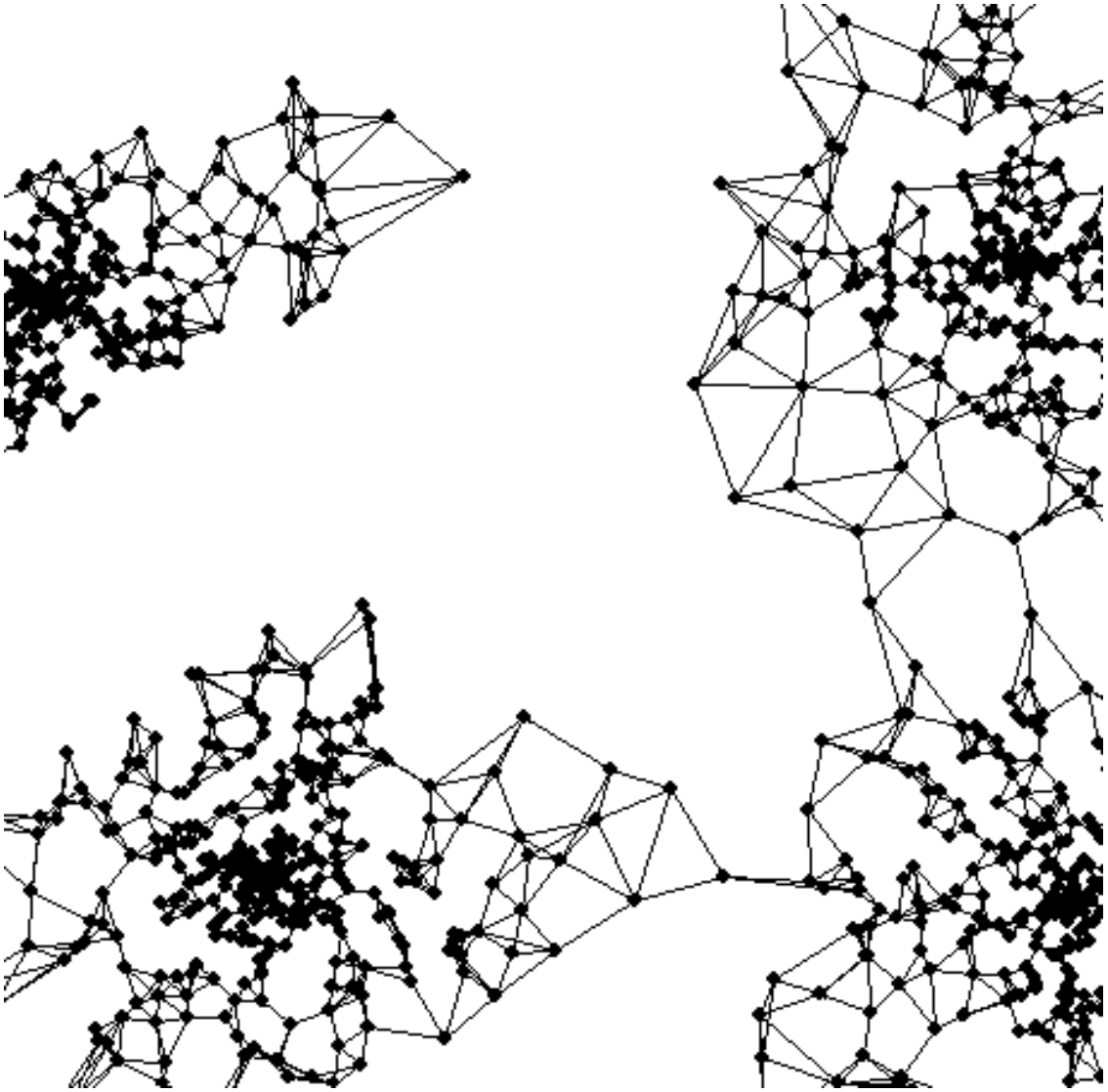
K nearest neighbor graph

k=3



K nearest neighbor graph

k=4



Bottleneck issues

What we have:

- No good way to setup *k* automatically
- It does not adapt to local density

What we want:

- Setup *k* automatically
- Value *k* should be small, preferably constant.
- Connected graph
- Constructing the graph efficiently

Neighborhood graphs

KNN:

- Parameter *k* needs to be setup experimentally
- Graph may not be connected

MST:

- Connected
- Corresponds to *k*=1 but constrained by connectivity

k-MST:

- MST repeated *k* times for remaining edges
- Average number of links equals to *kNN*
- Inherits connectivity of MST

Delanuay triangulation:

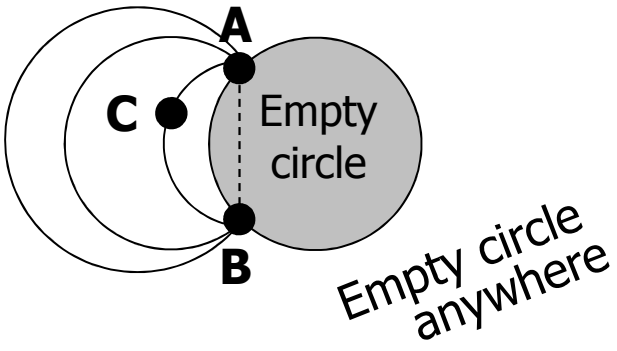
- Partition space by Voronoi diagram
- Connected
- High number of edges
- Exponential time complexity $O(n^{d/2+1})$

X-nearest neighbor graph
(XNN)

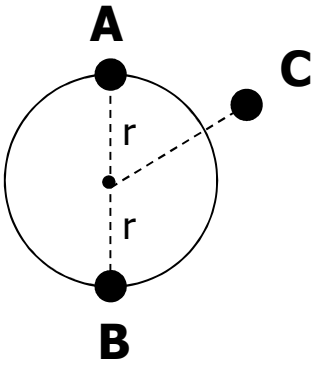
Neighbor rules

Basic definition:
A and B are neighbors,
if no other points within the circle.

Delaunay rule



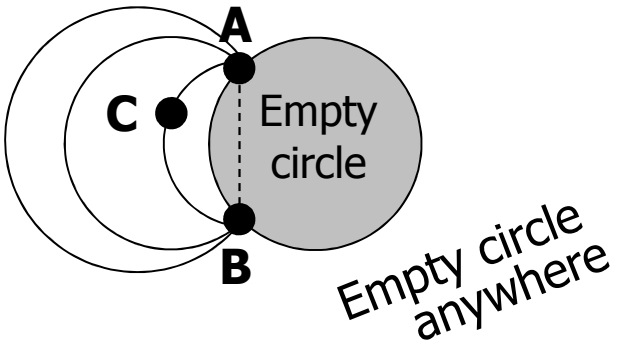
Gabriel rule



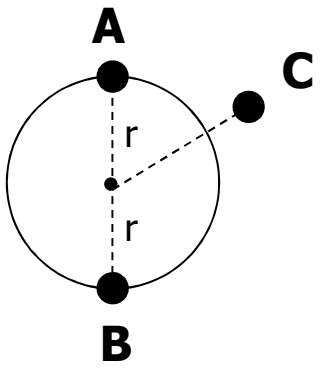
Neighbor rules

Basic definition:
A and B are neighbors,
if no other points within the circle.

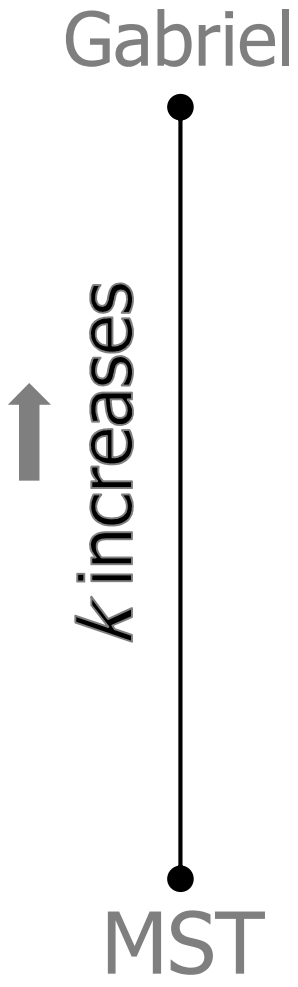
Delaunay rule



Gabriel rule



Three variants



1. Full XNN
2. Hierarchically-built XNN

3. k-limited XNN

Neighbor rules

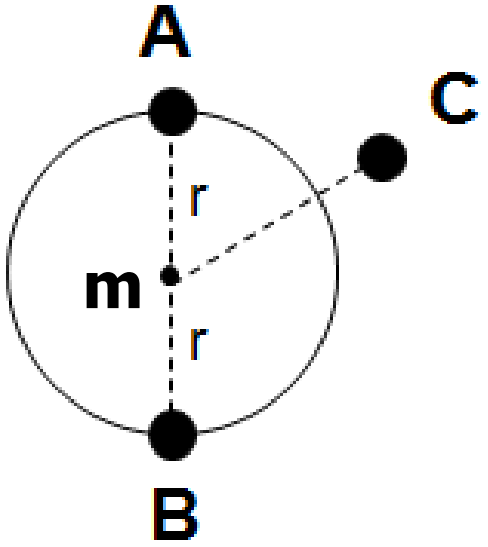
distances

Gabriel rule:

$$ab^2 < ac^2 + bc^2$$

Using midpoint:

$$ab < 2mc$$



C is the nearest point to the midpoint m

Neighbor rule
similarities

Assume similarity-to-distance conversion:

$d=1/(s+\varepsilon)$
where $s \in [0,1]$
$\varepsilon=0.001$

Neighbor rule:

$$ab^{-2} < ac^{-2} + bc^{-2} \Leftrightarrow \frac{ab^2 \cdot bc^2 + ab^2 \cdot ac^2}{ac^2 \cdot bc^2} > 1$$

Algorithm to create full XNN

Algorithm 1: Full XNN(data set) → XNN

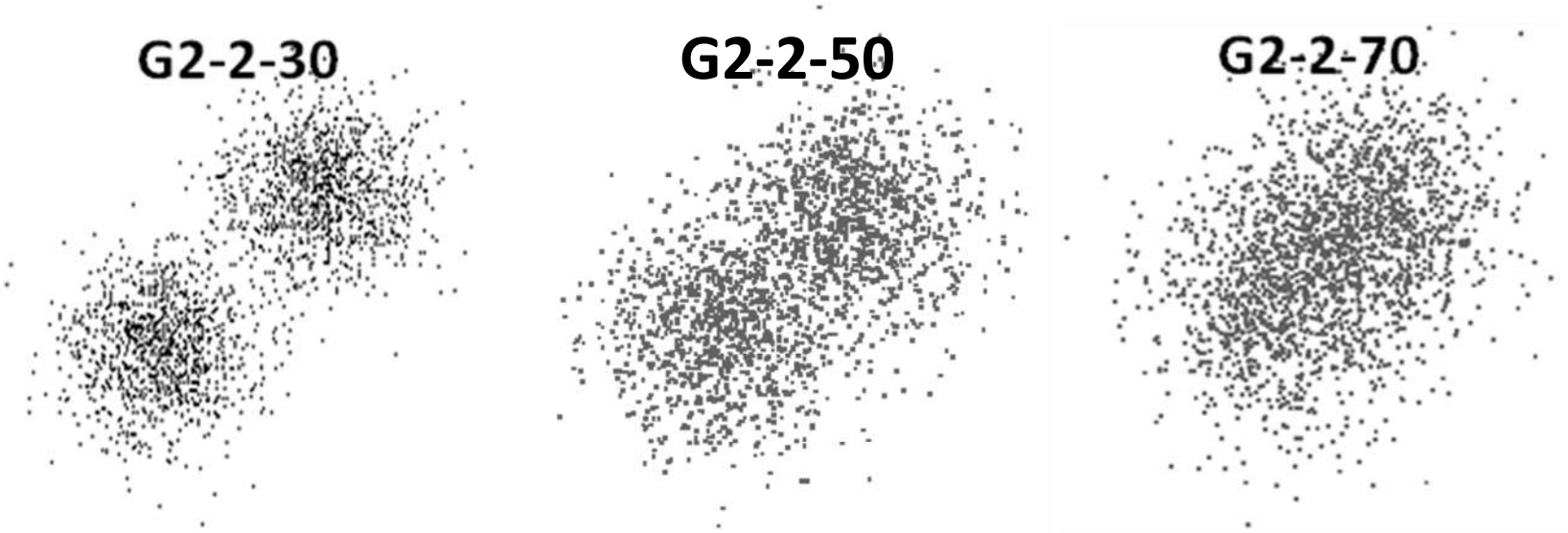
```
FOR i=1 to N-1 DO
   FOR j=i+1 to N DO
      Calculate midpoint m ← (xi+xj)/2
      x ← Find nearest point for m
      IF x== xi OR x== xj THEN
            Mark xi and xj as neighbors.
```

$O(N^3)$

How many neighbors…?

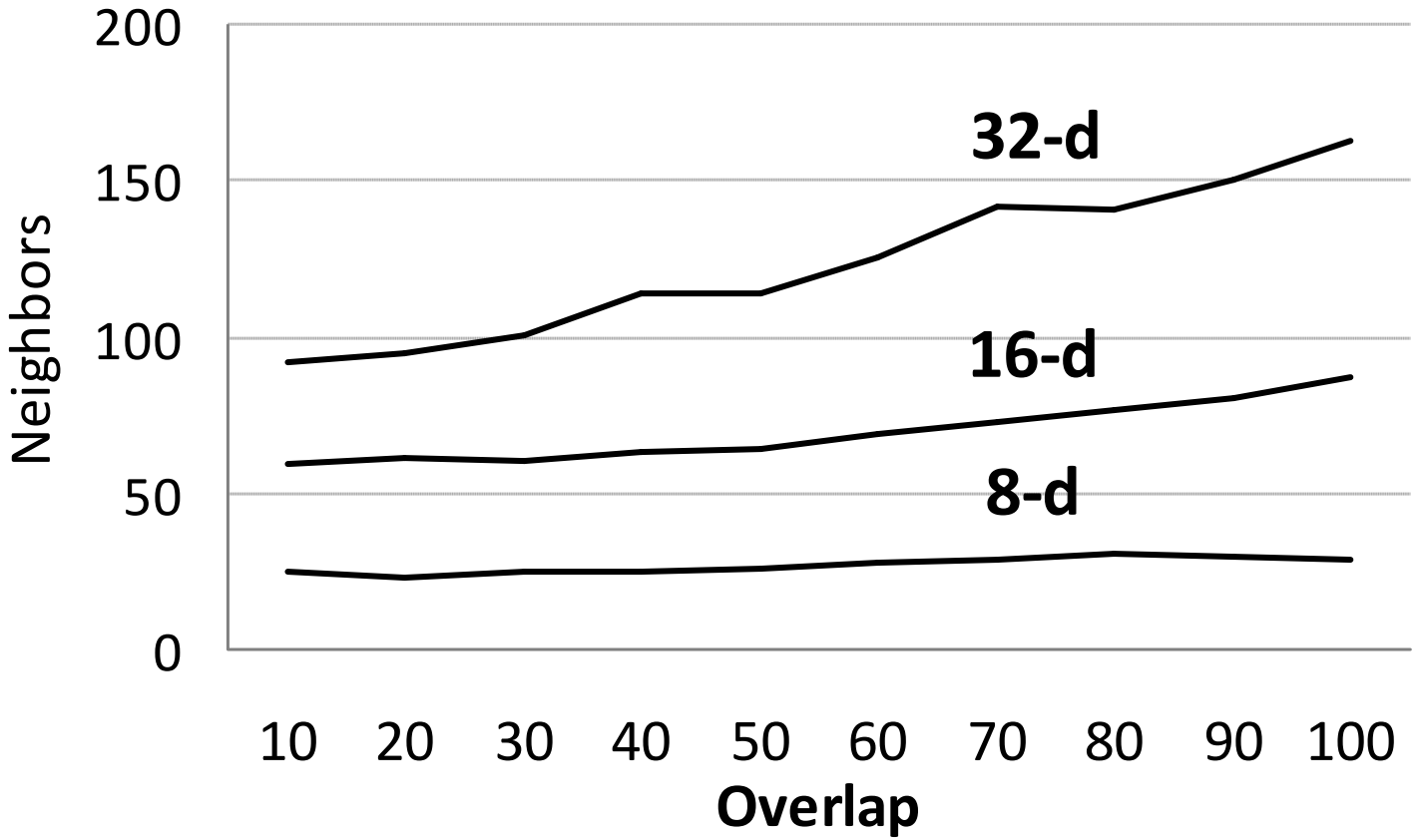
Experimental observations

- G2 data sets
- Two clusters of 500 points each
- Varying overlap (10%..90%)



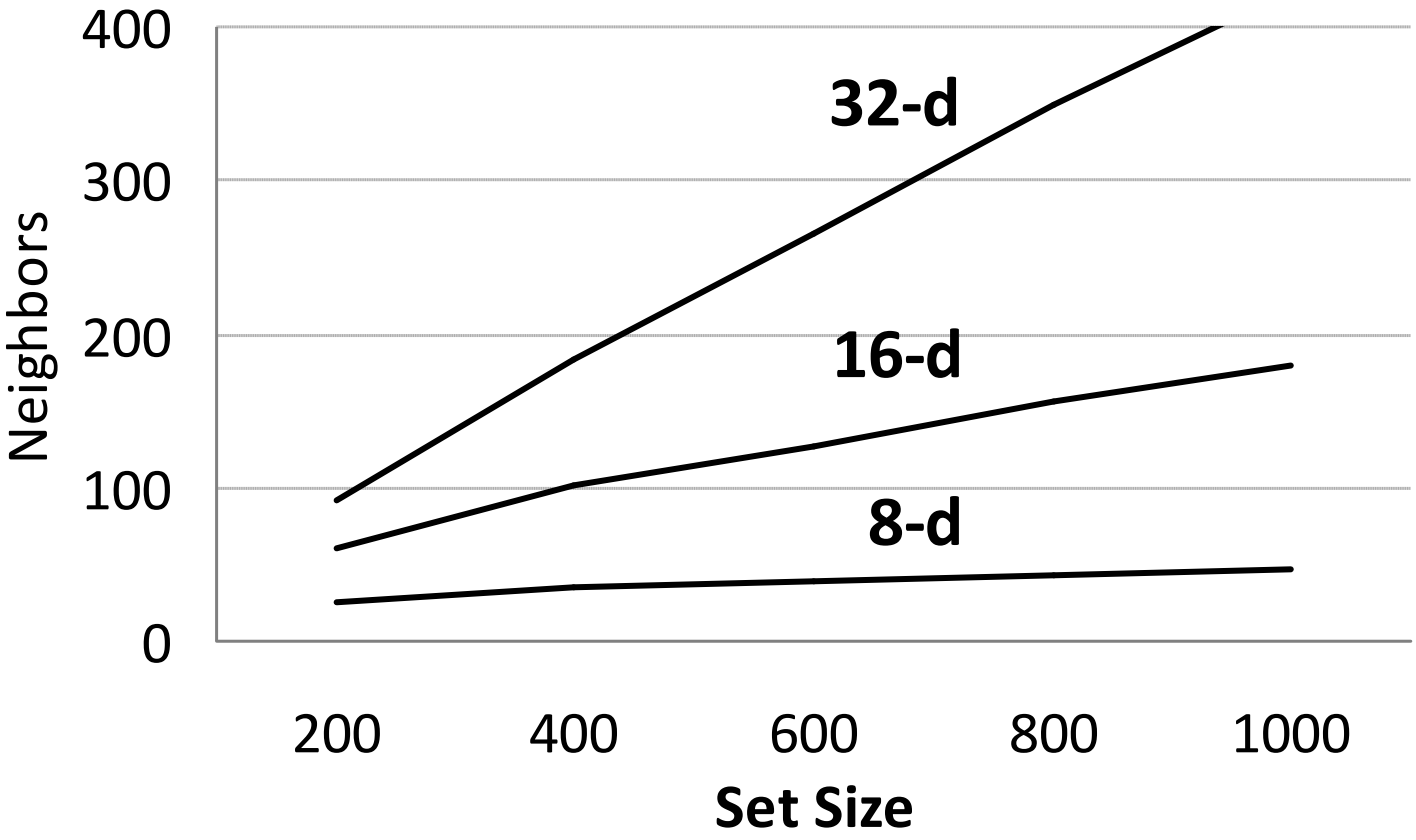
Effect of overlap

G2-32-xx



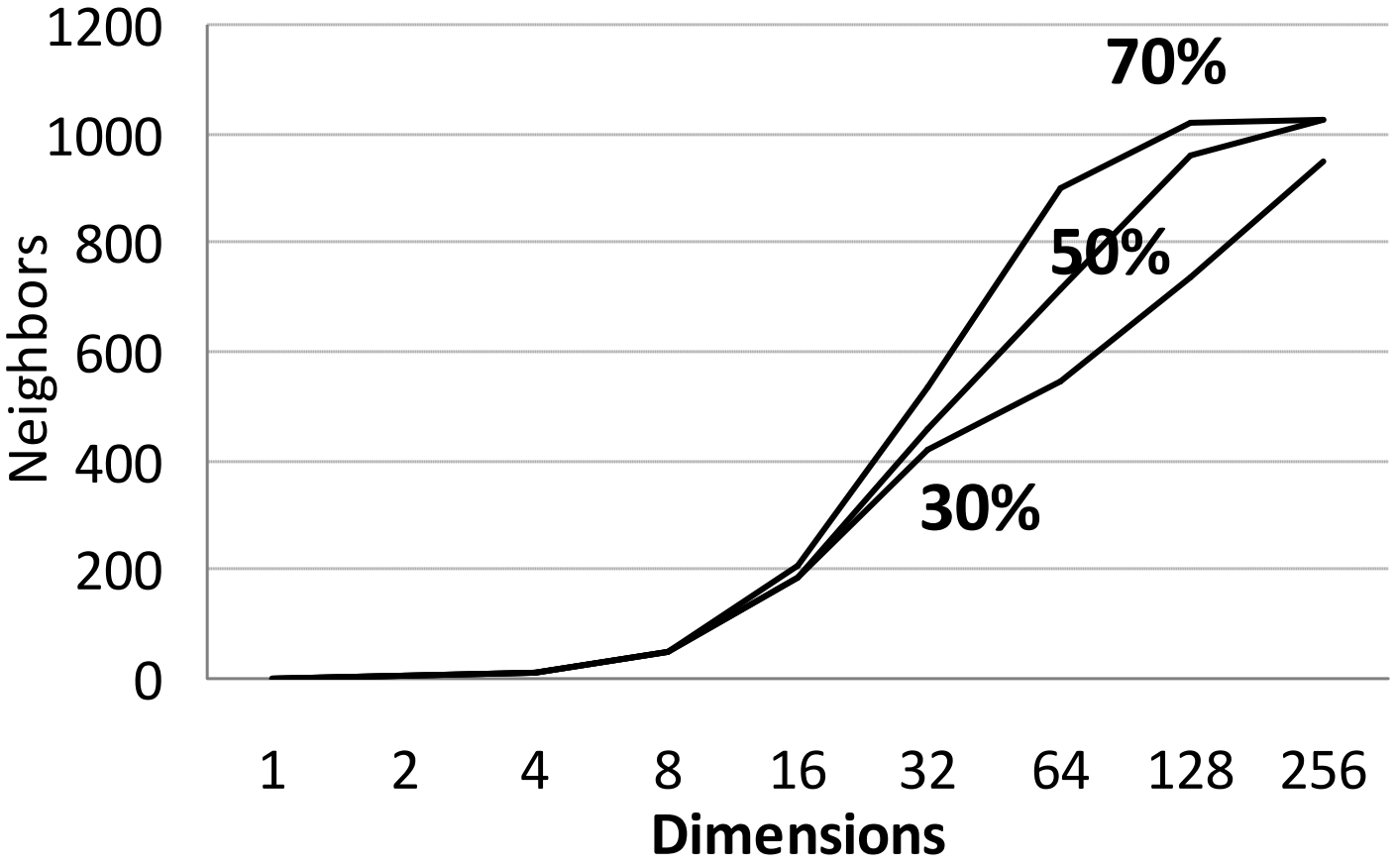
Effect of density (varying N)

G2-32-10



Effect of dimension

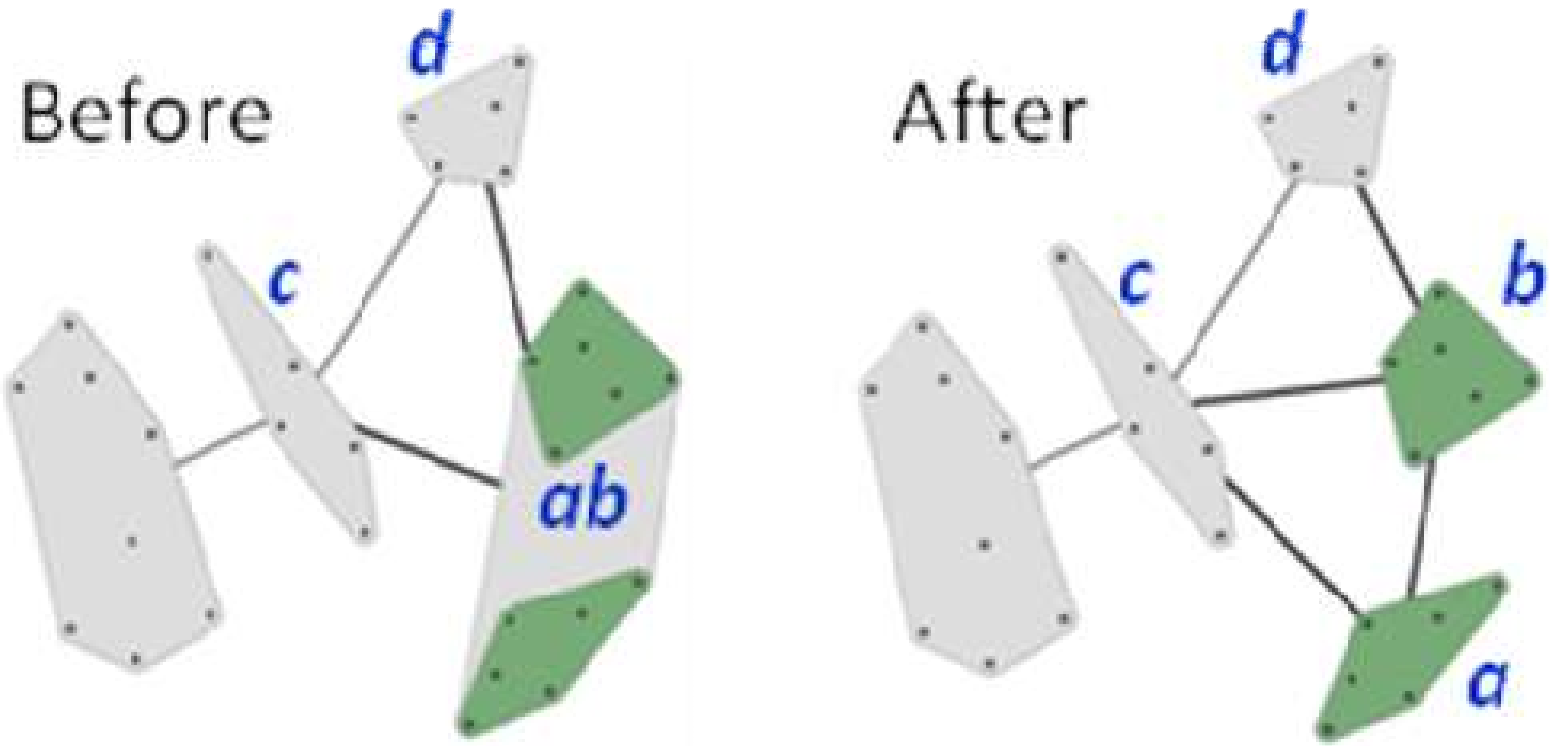
G2-xx-70



Hierarchical XNN

Algorithm 2: Hierarchical XNN(data) $\rightarrow$ XNN
Put all in one cluster;
XNN $\leftarrow \varnothing$;
WHILE |XNN| < N DO
    ab $\leftarrow$SelectLargestCluster(C);
    a, b $\leftarrow$ SplitCluster(*ab*);
    XNN $\leftarrow$ XNN $\cup$ (a,b)
    UpdateXNN;

Update after split

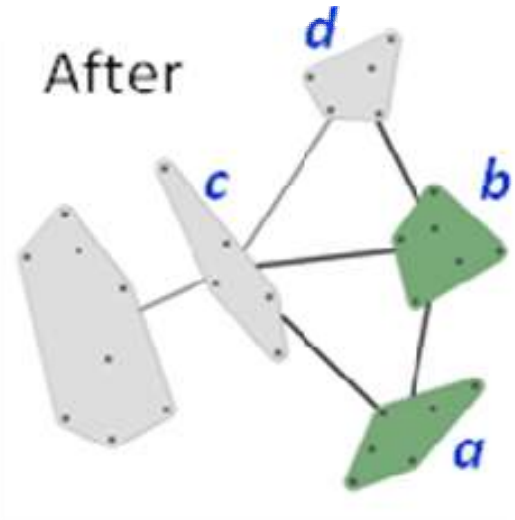


- Add link *ab* **always**
- Choose subset of {*ac*, *bc*, *ad*, *bd*}

Which links to choose?

(*ac* or *bc*) (*ad* or *bd*)

- Accept both $\Rightarrow$ complete graph
- Accept shorter $\Rightarrow$ spanning tree
- Neighbor rule $\Rightarrow$ hierarchical XNN



Time complexity:

$$X \cdot \sum_{i=1}^{N} i = X \cdot N^2$$

Number of links:

Still too high

K-limited XNN

- Select *ab*
- Select 2(*k*-1) closest among the rest
- *k* is global parameter (as in KNN)

$\Rightarrow$

Average number of links per point $\leq k$

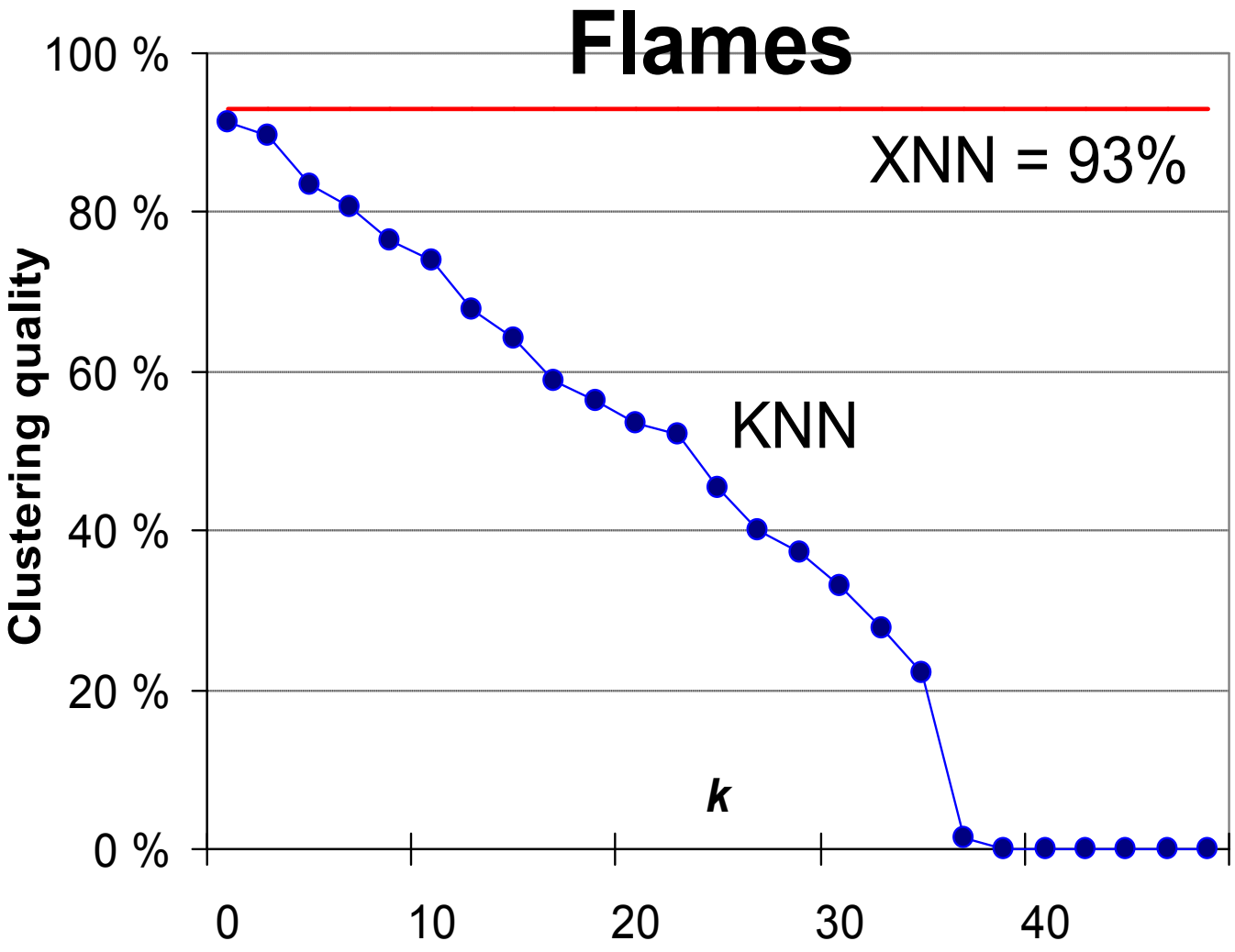
Number of neighbors

k=10

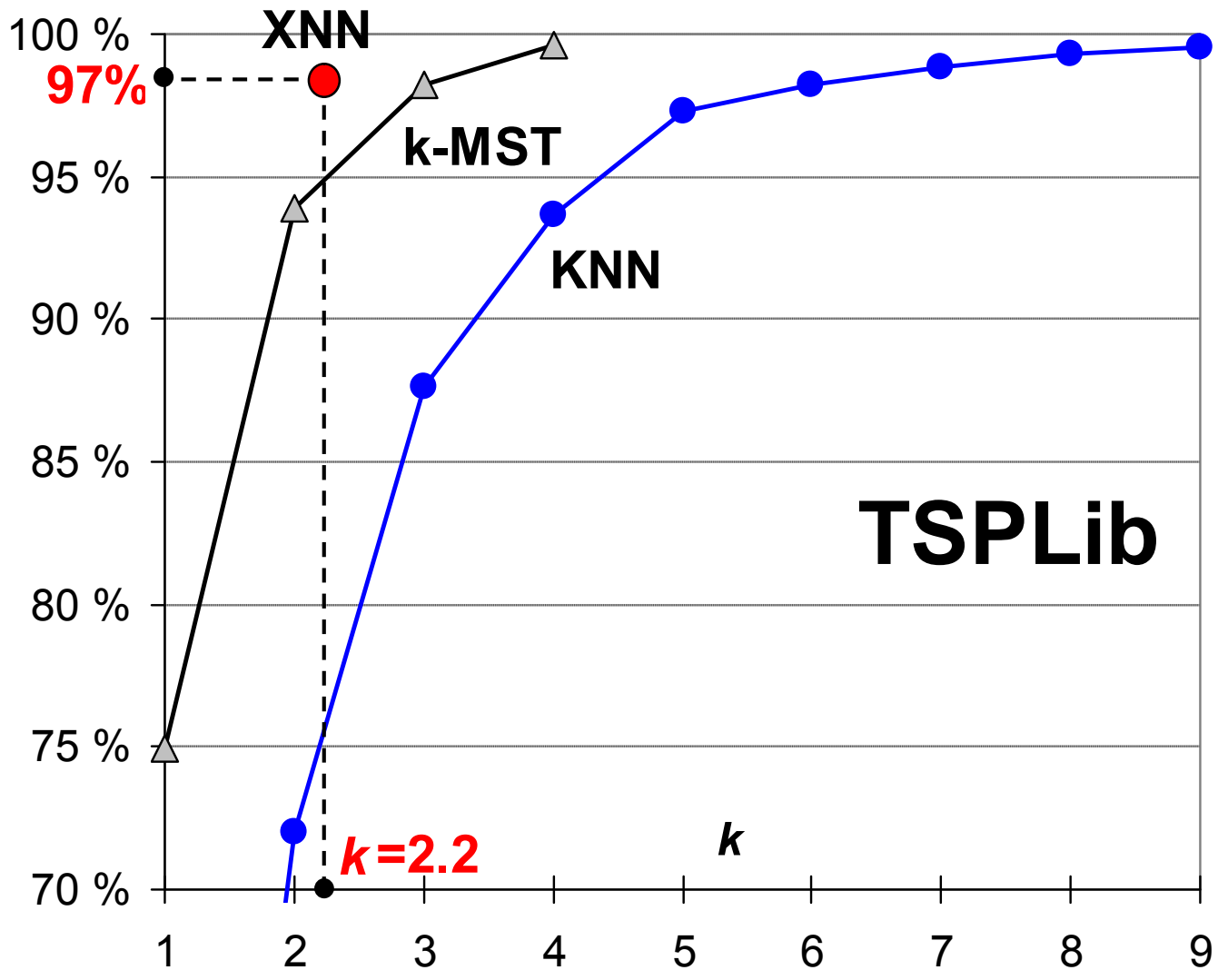
	Dataset	Dim	Full	Hierarchical	k-limited
16D	Bridge	16	68.8	48	6.5
3D	House	3	14.4	22	7.9
16D	Miss America	16	345	94	6.8
	Birch1	2	4.0	(3.4)	(3.4)
2D	Birch2	2	(3.7)	(3.4)	(3.4)
	Birch3	2	(3.9)	(3.4)	(3.3)
	S1	2	3.8	3.4	3.3
	S2	2	3.9	3.4	3.4
2D	S3	2	3.9	3.4	3.4
	S4	2	3.9	3.4	3.4

Applications examples

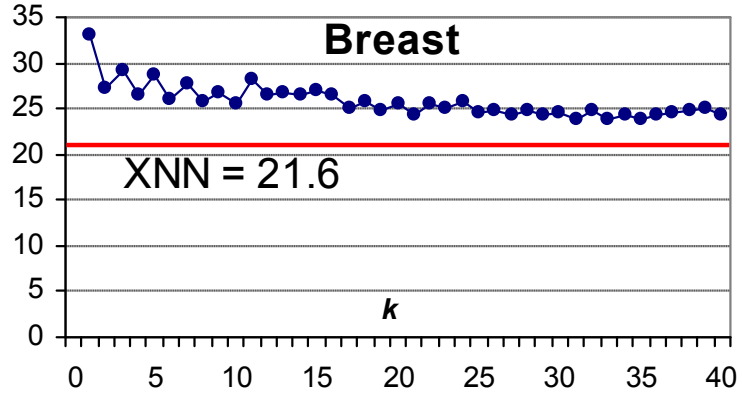
Path-based clustering

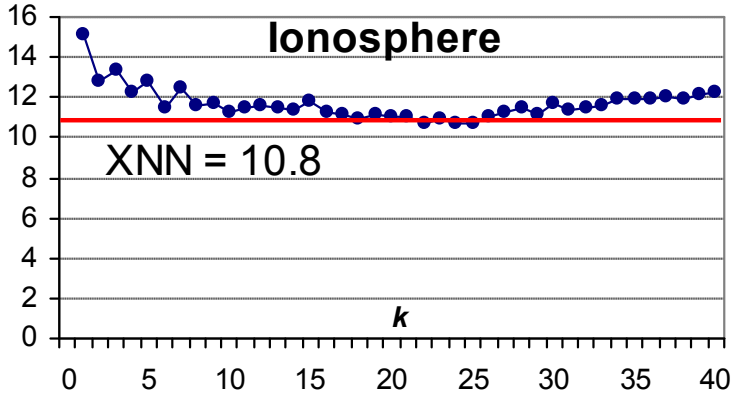


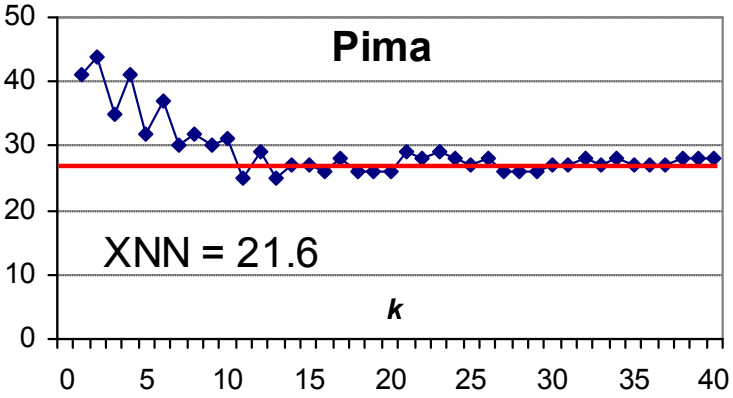
Travelling salesman problem

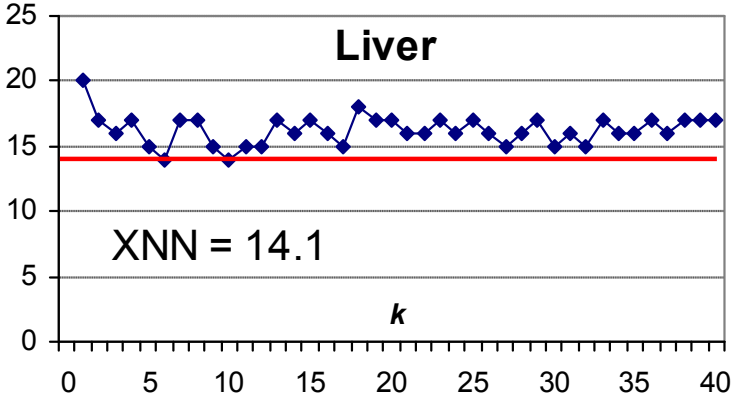


KNN classifier









Travelling salesman problem

Dataset	Points	Edges		Common	
	N	Total	Per node	Total	Per point
eil101	101	229	2.3	98	97%
a280	280	750	2.7	280	100%
RAT575	575	1213	2.1	552	96%
PR1002	1002	2060	2.0	957	96%
PR2392	2392	5127	2.1	2306	96%

Conclusions

- New neighborhood XNN
- Compromise between KNN and Gabriel
- It is connected
- Neighborhood size automatically selected