Weakest link clustering and convex geometry I

Boris Mirkin (National Research University Higher School of Economics, Russia)

DATA ANALYSIS AND MODELING LAB
Palacky University, Olomouc, Czech Republic

european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT

Weakest link clusters and convex geometries: an ordinal framework

Boris Mirkin

- Department of Data Analysis and Machine Intelligence, NRU HSE Moscow RF (partial support by Academic Foundation of NRU HSE 2009, 11-12, and LATNA NN HSE 2012-13 is acknowledged)
- Department of Computer Science, Birkbeck University of London, UK

Historic Intro to weakest link clusters

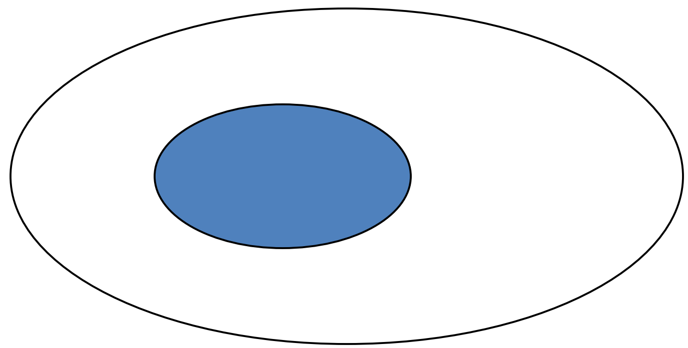
- **Joseph Mullat (1976** in Russian)
- **Ilya Muchnik and Co (1980-1990** in Russian) applied to organization structure, to consensus ordering etc.
- **Boris Mirkin, Ilya Muchnik (1997, 2002) –** some mathematics
- Similar work:
  - **Vladimir Batagelj (2002** and later – a special case, package **Pajek** for graph structure analysis)
  - others

Outline: 1st part, 2d part

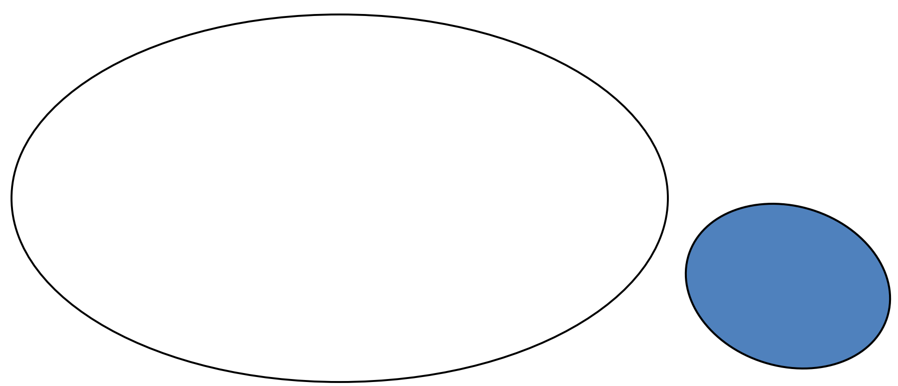
- Monotone linkage and tightness functions
- Shell layered clusters
- Serial partitioning algorithm
- Constrained shell layered clusters
- Monotone linkage and hereditary set-to-subset mapping
- Interior set system and Convex geometry
- Conclusion

Single cluster

- Core: today



- Anomalous: yesterday



Monotone linkage function $\pi(k,S)$

Inner: $k \in S$

- **in(k,S)**= $\sum_{j\in S} g_{kj}$ – incidence number [$G=(g_{kj})$ square graph incidence or rectangular context 1/0 matrix]
- **in2(k,S)**= $\sum_{j\in S} g_{kj} \sum_{t\in I} g_{jt}$ – second degree incidence number
- **c(k,S)**= $|\cup_{j\in S-k}\Gamma_j - \Gamma_k|$ - # complementary attributes/ control functions

Monotone increasing:

$$\pi(k,S) \leq \pi(k,S\cup T)$$

Monotone linkage function $\pi(k,S)$

Outer: $k \notin S$

- **L(k,S)**$=\min_{j\in S} d_{kj}$ single linkage

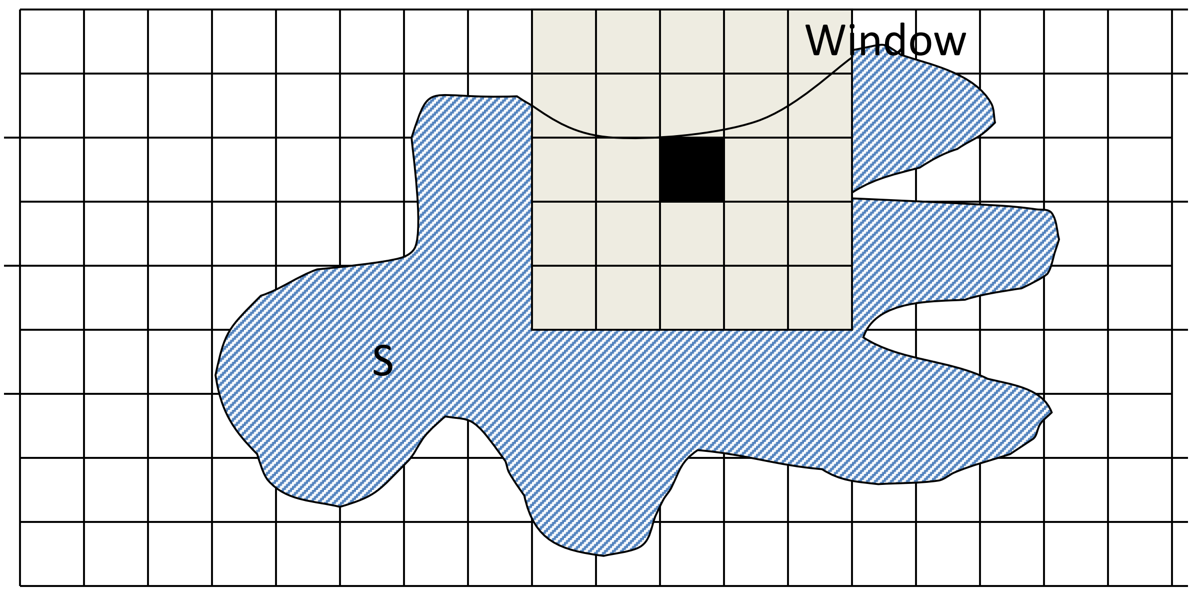
- **Ml(k,S)**$=\sum_{v\in V} \min_{j\in S}|x_{kv}-x_{jv}|/|V|$ - average minimum difference

Monotone decreasing:

$$\pi(k,S) \geq \pi(k,S\cup T)$$

Linkage function at an image

Linkage: count of pixels in $S \cap$ Window



Four monotonicity classes

Range Monotonicity	Inner	Outer
Growing	**I stick to**	
Falling		

Define **G(S)=F(I-S)**: dual inner- outer properties

Tightness function

- Given a monotone increasing **π(k,S),** its **Tightness (Weakest Link)** function

$$F_{\pi}(S) = \min_{k \in S} \pi(k,S)$$

Quasi-Convexity property

$$F_{\pi}(S \cup T) \geq \min\, [F_{\pi}(S), F_{\pi}(T)]$$

Tightness function and quasi-convexity

$$F(S \cup T) \geq \min\,[F(S), F(T)]$$

Given a quasi-convex set function **F(S)**, define

$$\pi_F(k,S) = \max_{k \in T \subseteq S} F(T)$$

which is monotone increasing

Theorem (A. Malishevski 1986, 1998).

Duality: $F_{\pi_F} = F$

Weakest link mechanism to generate all the quasi-convex functions: but HOW? No generator is known.

Weakest link core cluster

Given a monotone increasing **π(k,S),**

its **Weakest Link** cluster is **S** maximizing

$$F_\pi(S) = \min_{k \in S} \pi(k,S)$$

Not practical: Given a quasi-convex **F(S),** its maximizer is a Weakest Link cluster

Define $P_A(S)=1$ if $S=A$, $P_A(S)=0$, otherwise:

$P_A(S \cup T) \geq \min [P_A(S), P_A(T)]$. Yet HOW can one find the maximizer of P_A without enumerating all?

Shelled core cluster

- **F-Pattern S definition:**

  $$F(S) > F(S') \text{ if } S' \cap I\text{-}S \neq \emptyset$$

- **Patterns are chain nested**

- **If F is a tightness function, its minimum pattern is the largest maximizer of F**

Seriation algorithm for pattern finding I

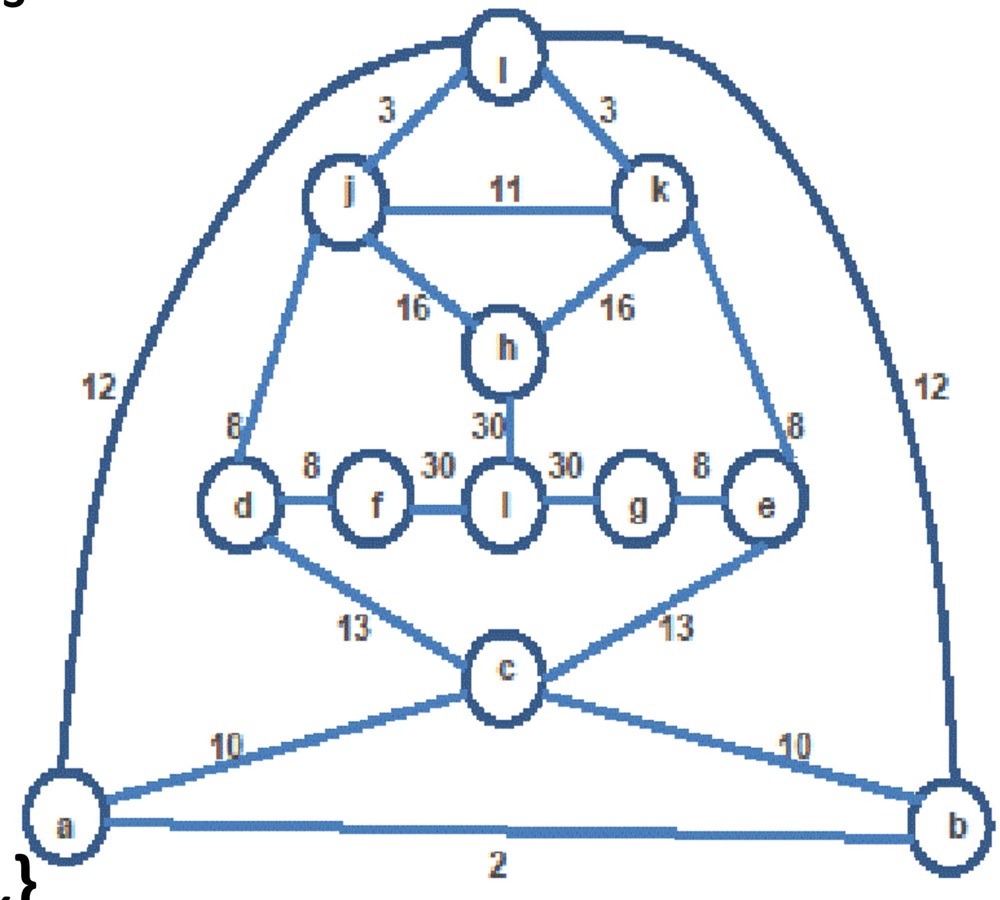
$M(S)=\{k: \pi(k,S)=\min_{j\in S} \pi(j,S)\}$

$T:=0;\ M:=\varnothing$

1. $T:=T+1$, $I:=I-M$

2. $S_T=M(I)$;
$M:=M(I);\ I:=I-M$

3. If $I=\varnothing$, end;
else go to 1



Output:

series $SS=\{S_1, S_2, \ldots, S_K\}$

PATTERN = Final_SS_Fragment_At_Max_π_Value

Seriation algorithm for pattern finding II

$\pi(k,S)=\sum_{j\in S} A(k,j)$

$S_1=M(l_1)=\{a,b\}$ at $\pi=24$

$S_2=M(l_2)=\{l\}$ at $\pi=6$

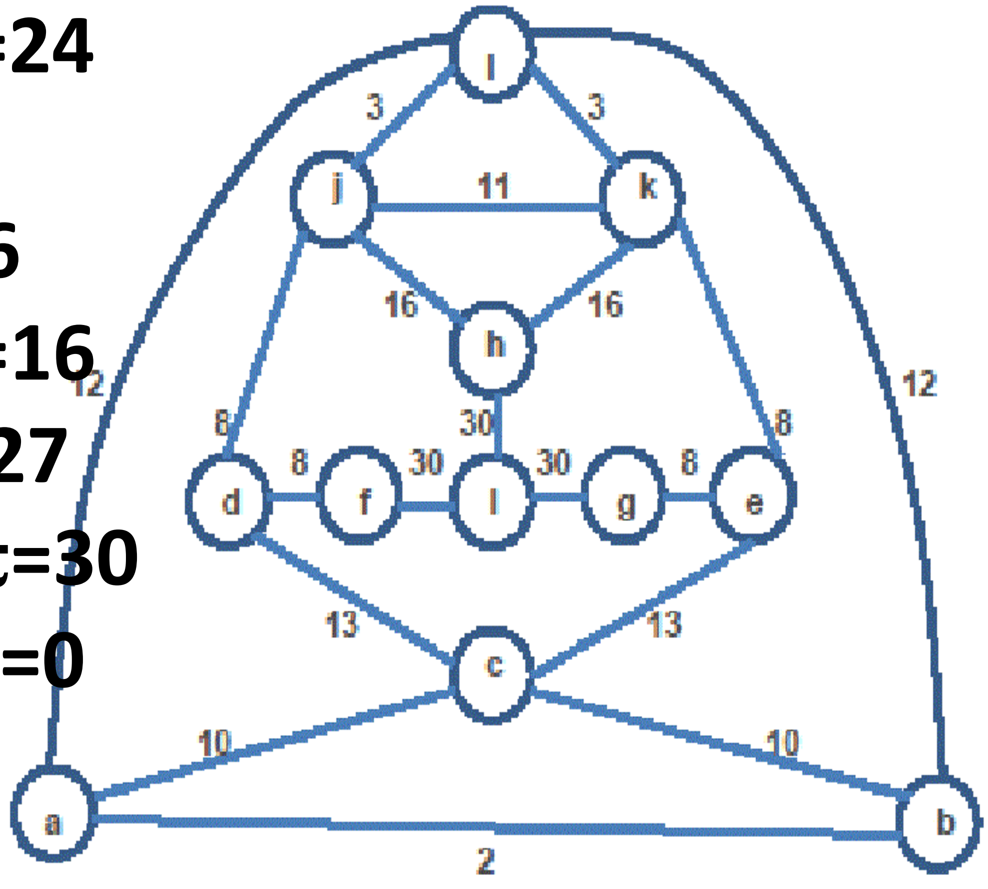
$S_3=M(l_3)=\{c\}$ at $\pi=26$

$S_4=M(l_4)=\{d,e\}$ at $\pi=16$

$S_5=M(l_5)=\{j,k\}$ at $\pi=27$

$S_6=M(l_6)=\{f,g,h\}$ at $\pi=30$

$S_7=M(l_7=\{l\})=\{i\}$ at $\pi=0$



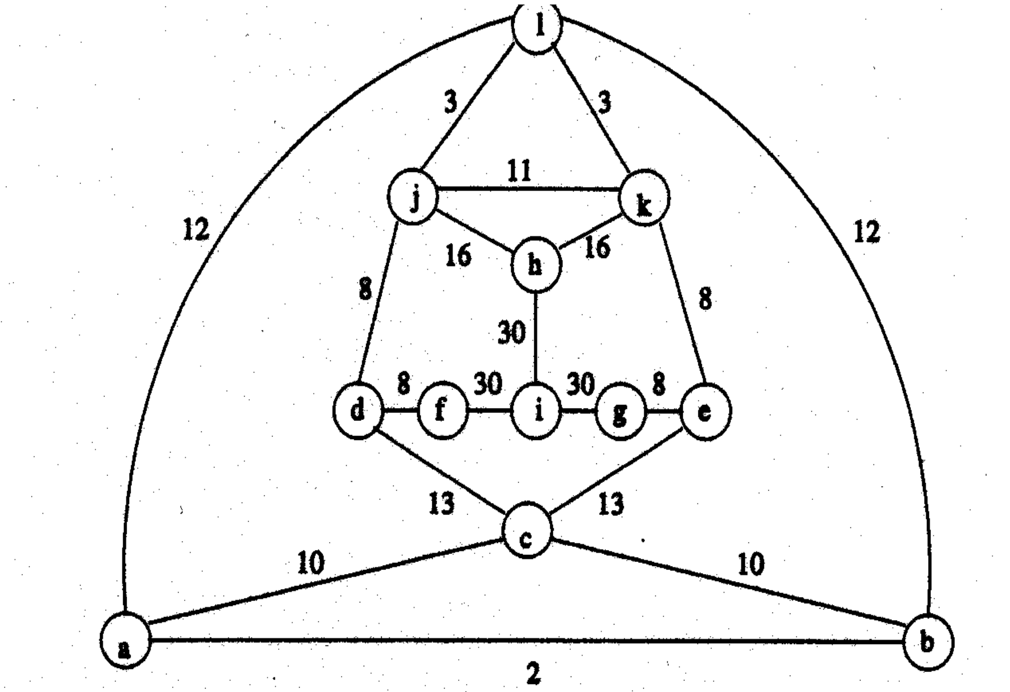
Shelled core cluster

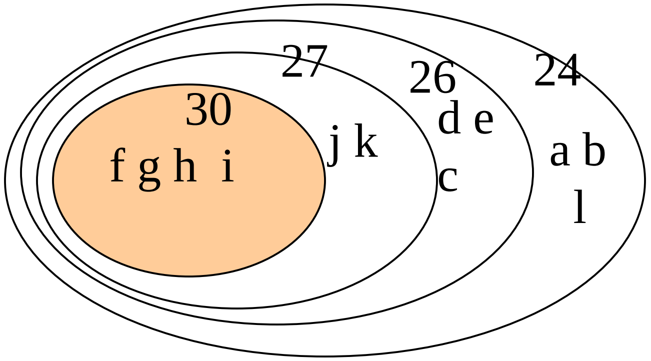
P1=fghi ^{30};   P2= fghijk ^{27};   P3=fghijkdec ^{26};   P4=I ^{24}

Structural model: shelled core

- Shelled core:

$(ab)^{24}(l)^{6}(c)\ ^{26}(de)^{16}(jk)^{27}(fgh)^{30}(i)^{0}$





Added constraints brought in by external forces

- Monotone linkage function $\pi(k,S)$
- Monotone external force linkage function $d(k,S)$

❖ Constraint $d(k,S) \leq u$ ("oversensitive" are not feasible) brings forth mapping

$$\psi_u(S)=\{\ k:\ d(k,S)\leq u\}$$ (assume not empty!)

❖ Tightness constrained:

$$F_{\pi u}(S)= \min_{k\in\psi u(S)} \pi(k,S)$$

Added constraints brought in by external forces

- Monotone linkage function **π(k,S)**
- Monotone external force linkage function **d(k,S)**

❖ $\varphi_u(S)=\{ k: d(k,S)\leq u\}$ **(assume not empty!)**

❖ **Tightness constrained:**

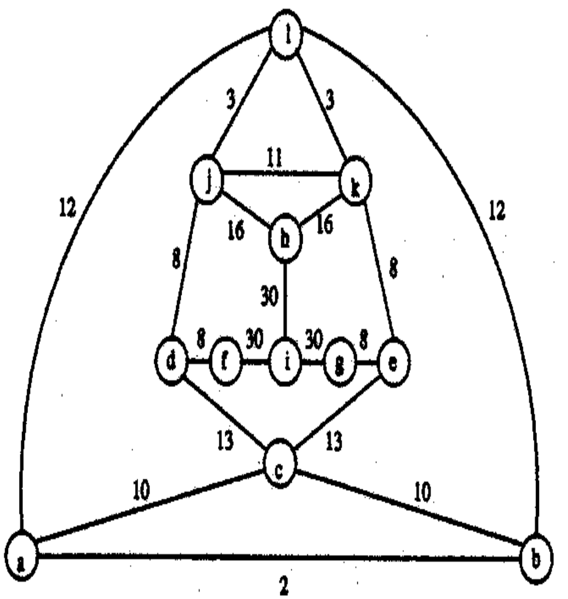
$$F_{\pi u}(S)= \min_{k\in\varphi u(S)} \pi(k,S)$$

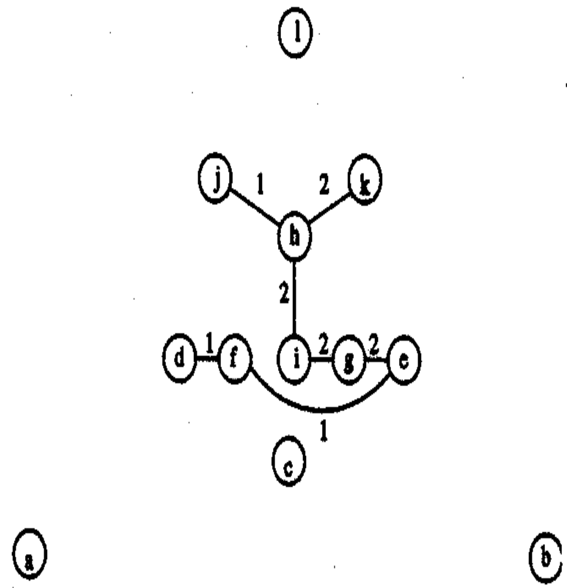
❖ Change **π(k,S):**

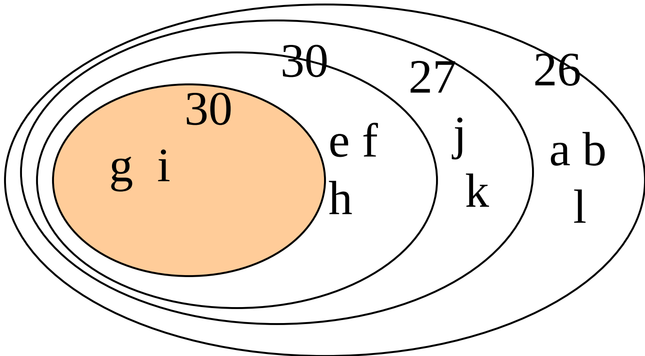
$$\pi'(k,S)=\begin{cases}\pi(k,S) & if \quad d(k,S)\leq u \\ \pi(k,S)+C & if \quad d(k,S)>u\end{cases}$$

❖ $F_{\pi u}(S) = F_{\pi'}(S)$

Added constraints brought in by external forces: $d(k,S)=\sum_{j\in S} h_{kj} \le 2$







- Shelled core under the constraint

Conclusion 1
Concepts considered

- Monotone linkage function: $I \times 2^I \Rightarrow R$
- Tightness function/Weakest link: $2^I \Rightarrow R$
- Quasi-convexity property
- Greedy-wise finding the shelled core
- Constraints defined by another monotone linkage function