# Induced Layered Clusters, Hereditary Mappings and Convex Geometries*

Boris Mirkin† Ilya Muchnik‡

March 13, 2001

**Abstract**

A method for structural clustering proposed by the authors is extended to the case when there are externally defined restrictions on the relations between sets and their elements. This framework appears to be related to order-theoretic concepts of the hereditary mappings and convex geometries, which enables us to give characterizations of those in terms of the monotone linkage functions.

**Key words:** layered cluster, monotone linkage, greedy optimization, convex geometry, hereditary mapping.

## 1 Introduction

In [1] we explored the concept of layered cluster on a set $I$ of elements whose interrelations are characterised by a set-to-element linkage function. In this paper we extend this analysis to the situations in which the set $I$ can be surrounded by other elements that may considerably affect the functioning of the elements of $I$. For example, a spatial protein folding is typically partitioned in a set of spatially separated substructures called domains. The concept of layered cluster can be applied to each of the domains separately. However, the domains may be spatially close and thus affect one another. To take into account the between-domain interaction, a modified form of the model proposed in [2] can be utilised. We assume that there are two linkage functions on $I$: the straight within-domain linkage function $\pi(i, H)$ of [1] and an induced linkage function $d(i, H)$ that reflects the sensitivity of an element to the external forces. The "oversensitive" elements should be excluded from the definition of the tightness function: for any subset $H$, its tightness, that is, the minimum of the straight linkages $\pi(i, H)$ $(i \in H)$, should be defined over only the non-sensitive elements, viz. those elements $i \in H$ satisfying the constraint $d(i, H) \leq u$ where $u$ is an appropriately chosen sensitivity threshold (the induced tightness).

In the remainder we prove that this restriction doesn't affect the properties of the tightness function and the induced layered cluster can be found by using a modified version of the greedy serial partitioning algorithm proposed in [1]. Then we show that the set of all possible induced patterns forms a convex geometry [3] and characterize the convex geometries in terms of the monotone linkage functions and related terms of the hereditary set-to-subset mappings [4].

## 2 Induced tightness function and its layered core

Let two linkage functions, $\pi(i, H)$ and $d(i, H)$, be defined for all $H \subseteq I$ and $i \in H$. Each is assumed to be monotone over sets $H$ so that any increase in $H$ may not decrease the linkage value. The second linkage function $d(i, H)$ will be referred to as the induced linkage function since it is considered

*The work was done in the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) at Rutgers University. The authors thank DIMACS' Director Dr. F. Roberts for his support. The authors are indebted to T. Fenner and G. Loizou for thorough discussions of the material and help in better shaping of it.

†School of Computer Science and Information Systems, Birkbeck College, London WC1E 7HX, UK, e-mail: mirkin@dcs.bbk.ac.uk

‡DIMACS, Rutgers University, Piscataway NJ, 08854-8018, USA, e-mail: muchnik@dimacs.rutgers.edu

reflecting the external forces. For example, if elements of $I$ are affected by a set $K$ of the "outer" elements and the intensity of interactions between $i \in I$ and $k \in K$ is scored by an index, $a_{ik}$, then the induced similarity, $s_{i'i''}$, between $i' \in I$ and $i'' \in I$ can be defined as the sum or maximum or any other function of $a_{i'k}$ and $a_{i''k}$ over all affecting $k \in K$. The induced linkage function $d(i, H)$ can be defined as the summary similarity $d(i, H) = \sum_{j \in H} s_{ij}$.

For example, let the weighted graphs in Figure 1 represent straight and induced similarities between the eleven nodes constituting set $I$. The graph on the left of Figure 1 defines $\pi(i, H)$ as the summary linkage function, $\pi(i, H) = \sum_{i \in H} s_{ij}$, and the graph on the right similarly defines the induced function $d(i, H)$. Thus defined $\pi(i, H)$ was considered in the example in [1] and its layered cluster was shown to consist of patterns $H_3 = \{f, g, h, i\}$, $H_2 = H_3 \cup \{j, k\}$, $H_1 = H_2 \cup \{c, d, e\}$, and $H_0 = I$.

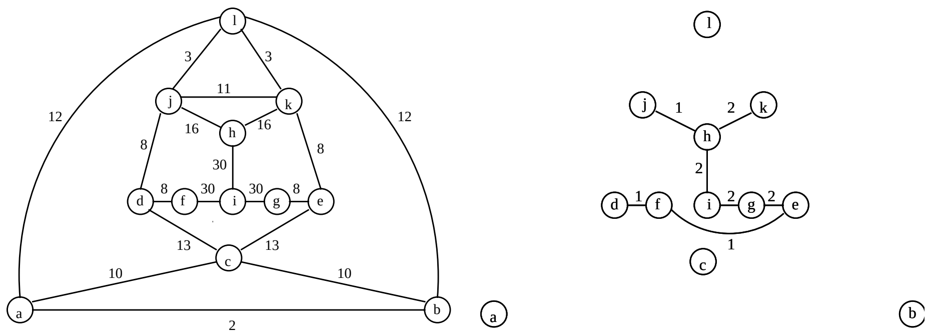


Figure 1: Two interrelation graphs between the circled entities.

For any $H \subseteq I$ and a prespecified threshold $u$, let us define its subset $\phi_u(H)$ as

$$\phi_u(H) = \{i \in H : d(i, H) \leq u\} \quad (1)$$

It is important for the follow-up mathematical analysis to have the mapping $\phi_u$ well-defined so that $\phi_u(H)$ is not empty if $H$ is not empty. Thus, we require the threshold to be large enough, $u \geq \max_H \min_i d(i, H)$, to provide for $\phi_u(H)$ being well-defined.

In the example under consideration, some induced linkages are as large as $d(h, I) = 5$, but the maximum value of its tightness function $F_d(H)$ is 2 and reached at $H = \{g, h, i, k, l\}$.

The induced tightness function can be defined then as

$$F_{\pi,d}(H) = \min_{i \in \phi_u(H)} \pi(i, H). \quad (2)$$

When $u$ is large enough so that $\phi_u(H) = H$ for all $H \subseteq I$, $F_{\pi,d}(H)$ is the straight tightness function $F_\pi(H)$ as defined in [1].

To extend the concepts and results of [1] to the case under consideration, let us introduce a straight linkage function whose action is equivalent to that of the induced one. This new linkage function, $\pi_d$, differs from $\pi$ by a high penalty imposed each time when $i$ may not interact within $H$ under the induced linkages, that is, when $d(i, H) > u$:

$$\pi_d(i, H) = \begin{cases} \pi(i, H), & if \quad d(i, H) \leq u \\ \pi(i, H) + c, & if \quad d(i, H) > u \end{cases} \quad (3)$$

Here $c$ is a sufficiently large number defined for instance as $c = \max_{i,H} \pi(i, H) + 1$.

**Assertion 1** *If $\pi(i, H)$ and $d(i, H)$ are monotone linkage functions and $\phi_u(H)$ is well-defined, then $\pi_d(i, H)$ is also monotone and its tightness function coincides with the induced tightness function, that is, $F_{\pi_d}(H) = F_{\pi,d}(H)$ for any $H \subseteq I$.*

**Proof:** Let us prove that $\pi_d(i, H)$ is monotone; that is, $\pi_d(i, H) \le \pi_d(i, H \cup G)$ for all $H, G \subseteq I$ such that $i \in H$. Logically, there can be four cases: (a) $d(i, H) \le u$ and $d(i, H \cup G) \le u$, (b) $d(i, H) \le u$ and $d(i, H \cup G) > u$, (c) $d(i, H) > u$ and $d(i, H \cup G) \le u$, (d) $d(i, H) > u$ and $d(i, H \cup G) > u$. In the cases (a), (b), and (d), monotonicity of $\pi_d$ obviously follows from the monotonicity of $\pi$, and case (c) is impossible because of the monotonicity of $d$.

The equality $F_{\pi_d}(H) = F_{\pi,d}(H)$ obviously follows from (3) when mapping $\phi_u(H)$ is well-defined. □

An induced pattern is defined as an $H \subseteq I$ such that $F_{\pi,d}(H') < F_{\pi,d}(H)$ for any $H'$ satisfying the condition $H' \cap (I - H) \neq \emptyset$. The Assertion 1 implies that the set of induced patterns is chain-nested and forms a layered cluster [1] which will be referred to as the induced layered cluster.

The algorithm of serial partitioning from [1] finds the induced layered cluster with mapping

$$md(H) = \{i \in \phi_u(H) : \pi(i, H) = F_{\pi,d}(H)\}$$

The algorithm iteratively computes subsets $I_{t+1} = I_t - md(I_t)$ and their tightness values $F_{\pi,d}(I_t)$, beginning with $I_0 = I$. The induced layered cluster consists of those subsets $I_t$ that correspond to recursively found maximal values of $F_{\pi,d}(I_t)$.

Let us consider the linkage functions $\pi$ and $d$ defined by the graphs in Figure 1. To guarantee that all the subsets $\phi(H)$ are nonempty, the threshold is set at $u = 2$ in (2).

Since the elements $a, b, c, l$ have zero $d$-linkages, they are permitted in (2) and, thus, constitute $md(I_t)$ removed at the two first steps, as in the nonrestricted problem [1]. After they are removed, the only entities satisfying condition $d(i, I - \{a, b, c, l\}) \le 2$ are $d, f, j, k$ of which $d$ has the smallest linkage, $\pi(d, I - \{a, b, c, l\}) = 16$. With $d$ removed, entities $j, k, f$ become admissible in (2) and $j$ brings the minimum linkage 27. Of $f$ and $k$ remaining after removing $j$, the minimum linkage, 24, is at $k$. In the remaining set $H = \{f, e, g, i, h\}$, only $f$ and $h$ are $d$-admissible, and both are removed as having the minimum linkage, 30, at this step. In thus found set $\{e, g, i\}$, $e$ is the obvious leader with the minimum linkage equal to 8, which leaves us with $g$ and $i$ linked together at 30. The resulting serial partition:

$$(ab)^{24}(l)^{6}(c)^{26}(d)^{16}(j)^{27}(k)^{24}(fh)^{30}(e)^{8}(gi)^{30}$$

where sets $md(I_t)$ are in the round brackets and corresponding values of $F_{\pi,d}$ are the powers.

The maximum subset corresponding to maximum of $F_{\pi,d} = 30$, is $H_3^d = \{e, f, g, h, i\}$; this is added by $k, j$ to form the next pattern, $H_2^d$ corresponding to the next maximum, 27. Further adding $d$ and $c$ to this makes $H_1^d$ corresponding to the next maximum 26. The induced core, $H_3^d$, differs from $H_3$ found in the non-restricted problem by $e$ moved in from $H_1$. In principle, the solution could be changed more drastically, if the $d$ structure differed from that of $\pi$ more significantly.

## 3 Hereditary mappings and monotone linkages

Let us consider a set-to-subset mapping, $\phi$, defined for any $H \subseteq I$ in such a way that $\phi(H) \subseteq H$. The mapping $\phi$ can be interpreted as a selection rule so that $\phi(H)$ is the set of selected entities in $H$ and $\psi(H) = H - \phi(H)$ the set of rejected entities in $H$.

Let us assume that the rejection mapping is isotone so that the set of rejected entities may only increase when $H$ increases. The isotonicity property can be expressed by the condition

$$\psi(H) \cup \psi(G) \subseteq \psi(H \cup G) \tag{4}$$

that holds for any $H, G \subseteq I$.

If the rejection mapping $\psi$ is isotone, the selection mapping $\phi$ must satisfy the dual condition

$$\phi(H \cup G) \subseteq \phi(H) \cup \phi(G) \tag{5}$$

or its equivalent

$$\phi(H \cup G) \cap H \subseteq \phi(H). \tag{6}$$

Condition (6) is well-known in the theory of social choice; a mapping $\phi$ is referred to as hereditary if it satisfies the condition (6) [4].

The following statement follows from the formulas above.

**Assertion 2** *Mapping $\phi(H)$ is hereditary if and only if its dual, $\psi(H) = H - \phi(H)$, is isotone.*

The concepts of monotone linkage function and hereditary mapping are equivalent in the following sense. A monotone linkage function, $d(i, H)$, defines a set of set-to-subset mappings as follows. For any real $u$,

$$\phi_u(H) = \{i \in H : d(i, H) \leq u\} \text{ and } \psi_u(H) = \{i \in H : d(i, H) > u\} \quad (7)$$

**Assertion 3** *If the linkage function $d(i, H)$ is monotone, then the mapping $\phi_u(H)$ is hereditary and $\psi_u(H)$ is isotone. Moreover, any hereditary mapping can be presented as $\phi_u$ in (7) for some monotone linkage function $d$.*

**Proof:** Let us prove that, for any threshold $u$, $\phi_u$ defined by (7) is hereditary if $u(i, H)$ is monotone. Let $H \subset G$ and, for an $i \in H$, $d(i, G) \leq u$. Then $d(i, H) \leq u$ because of the monotonicity, which proves that $\phi_u$ is hereditary.

Reversely, let us consider a hereditary mapping $\phi(H)$ and define a linkage function, $d_\phi$ by the condition that $d_\phi(i, H) = 0$ for $i \in \phi(H)$ and $d_\phi(i, H) = 2$ for $i \in H - \phi(H)$. Then, obviously, $\phi(H) = \{i \in H : d_\phi(i, H) \leq 1\}$. Let us assume $d_\phi(i, G) = 0$ for some $H \subset G$ and $i \in H$. To prove that $d_\phi$ is monotone, we need to prove that $d_\phi(i, H) = 0$ too. By the definition of $d_\phi$, $i \in \phi(G)$. Therefore, $i \in \phi(H)$ or, equivalently, $d_\phi(i, H) = 0$, because $\phi$ is hereditary. For different thresholds the mappings defined by $d_\phi$ are also hereditary since they either coincide with $\phi$ or are empty, $\phi_u(H) = \emptyset$ (for the negative thresholds $u$), or complete, $\phi_u(H) = H$ (for the thresholds greater than 2). □

Obviously, $u' > u$ implies $\phi_u \subseteq \phi_{u'}$. Reversely, a family of hereditary mappings $\phi_u$ $(u \in U)$, where $U$ is a set of numbers, which satisfies this property, corresponds to a monotone linkage function whose mappings (7) form this family. The proof is analoguous to that of Assertion 3.

# 4 The induced patterns and interior set system

Given a monotone linkage function $d(i, H)$ and a threshold $u$ such that the mapping $\phi_u$ is well defined, let us consider the set $P(d, u)$ of all possible induced patterns: an $H \subseteq I$ belongs to $P(d, u)$ if and only if $H$ is an induced pattern of $F_{\pi,d}$ for some monotone linkage function $\pi(i, H)$.

It appears, the set of all induced patterns $P(d, u)$ can be greedily found as follows.

Given a mapping $\phi(S)$, let us define a set system, $C(\phi)$, by the following recursive condition: (a) $I \in C(\phi)$, (b) if $H \in C(\phi)$ then any $G$ such that $H - \phi(H) \subseteq G \subseteq H$ is also in $C(\phi)$. Thus created a system $C(\phi)$ will be referred to as an interior set system. The intuition behind this concept can be related to the linkage-based definition of $\phi$: if $\phi(H) = \{i : d(i, H) \leq u\}$ then $\phi(H)$ can be considered the set of extreme elements of $H$ so that the removal of any of them does not destroy the interior of $H$.

**Assertion 4** *If $\phi_u(H)$ is well-defined, then $P(d, u) = C(\phi_u)$.*

**Proof:** Let us prove that, for any $\pi(i, H)$, its induced patterns belong to $C(\phi_u)$. Indeed, $I \in C(\phi_u)$ and $H - md(H) \in C(\phi_u)$ if $H \in C(\phi_u)$, because $md(H) \subseteq \phi_u(H)$. Therefore, in the serial partitioning algorithm, all sets $I_t$ (thus, all induced patterns) belong to $C(\phi_u)$. Conversely, any subset $S \in C(\phi_u)$ is the only maximizer of its characteristic function, $f_S(H)$ that is equal to 0 for all $H \neq S$ and 1 when $H = S$, which is a tightness function [1]. □

## 5 Interior set and pattern systems as convex geometries

A set of subsets, $L \subseteq 2^I$, forms a convex geometry if it satisfies the following conditions:

L1. Sets $\emptyset$, $I$ belong to $L$;

L2. If $A \in L$ and $B \in L$, then $A \cap B \in L$;

L3. If $A \in L$ and $A \neq I$, then there exists $i \in I - A$ such that $A + i \in L$.

The symbol "+" denotes adding an element to a set.

This concept was introduced in [3] and studied in [5], [6].

Let us prove one more property of the convex geometries, that is a converse to L3:

L4. If $A \in L$ and $A \neq \emptyset$ then $A - i \in L$ for some $i \in A$.

Indeed, if $A \in L$ is a singleton, the statement is trivial because of L1. Let $A$ has more than one element. Since $\emptyset \in L$, then, by L3, $\emptyset + i_1 = i_1 \in L$ for some $i_1 \in I$. Continuing this way, we can find a sequence $s = i_1 i_2 ... i_{|I|}$ such that all its starting fragments, $S_k = \{i_1, .., i_k\}$, belong to $L$ ($k = 1, ..., |I|$). Let $k \geq 1$ be the first index such that $A \subseteq S_k$. Then $A - i_k = A \cap S_{k-1} \in L$ by L2, which proves L4.

**Assertion 5** *A set system $L$ is a convex geometry if and only if it is the interior set system of a hereditary well-defined mapping $\phi$.*

Proof: Let us prove that the interior set system $L = C(\phi)$ of a well-defined hereditary mapping $\phi$ satisfies L1-L3. Indeed, $I \in L$ by definition and $\emptyset \in L$ by construction, which proves L1. To prove L2 and L3, let us consider, for any $A \in L$, a chain nested system $A_1, ..., A_m \in L$ such that $A \subseteq A_1$, $A_m = I$ and $A_{i-1} = A_i - \phi(A_i)$ for all $i = 2, ..., m$, which exists according to the definition of $C(\phi)$. This system allows us to build a chain of sets in $L$, $A$, $A + i_1$, $A + i_1 + i_2$,..., $I$ such that all the sets $A_1$,..., $A_m$ belong to the chain. This immediately proves L3. To prove that $A \cap B \in L$ in the nontrivial case when neither of the sets is part of the other, let as consider the chain above, $A$, $A + i_1$, $A + i_1 + i_2$,..., $I$ and such $k \geq 1$ that $B \subseteq A \cup \{i_1, ..., i_k\}$ but $B \not\subseteq A \cup \{i_1, ..., i_{k-1}\}$. Obviously, $i_k \in B$, by the definition of $k$, and $i_k \in \phi(A \cup \{i_1, ..., i_k\})$, by the construction of the chain. Thus, $i_k \in \phi(B)$ according to the hereditary property and $B - i_k \in L$. Continuing the process of clearing $B$ from the elements that are absent in $A$, within $L$, eventually leads to $A \cap B \in L$, which proves L2.

Reversely, for $L$ being a convex geometry, let us define a mapping, $\phi$, that generates $L$ as its interior set system, and prove that it is well-defined and hereditary. It follows from L2 that, for any $A \subseteq I$, its closure in $L$, $L(A)$, can be uniquely defined as the intersection of all those $B \in L$ that include $A$. For any $B \in L$ let us define its extreme, $E(B) = \{i \in B : B - i \in L\}$. The mapping $\phi$ is defined then by $\phi(A) = E(L(A))$. The fact that $E(B)$ is nonempty for any nonempty $B$ follows from L4 above. Assume $i \in E(L(A))$ so that $L(A) - i \in L$. Then $A \not\subseteq L(A) - i$, because $L(A)$ is the smallest element in $L$ including $A$. This implies that $i \in A$ and proves $\phi(A) \subseteq A$.

Let us take now $B \subseteq A \subseteq I$. Obviously, $L(B) \subseteq L(A)$. If $L(B) = L(A)$, then $\phi(B) = \phi(A)$ and $\phi(B) = \phi(A) \cap B$, the hereditary property. If $L(B) \subseteq L(A) - E(L(A))$, then $B \cap \phi(A) = \emptyset$, again implying the heredity. The option remaining, $L(A) - E(L(A)) \subset L(B) \subseteq L(A)$ implies $\phi(A) \cap L(B) \subseteq \phi(B)$, which completes the proof that $\phi$ is hereditary.

To prove that $L = C(\phi)$, we use the fact that $A \in L$ if and only if there exists an ordering, $s = i_1 ... i_{|I|}$ of $I$ such that all its starting sets $A_k = \{i_1, ..., i_k\}$, $k = 1, ..., |I|$ belong to $L$, $A = A_k$ for some $k$ and $i_j \in E(A_j)$ for all $j = 1, ..., |I|$. For any $A \in L$ such a system is easy to build because of L3 and L4. On the other hand, $C(\phi)$ obviously is part of $L$. □

Assertions 4 and 5 prove that the concept of convex geometry is equivalent to the concept of set of all induced patterns $P(d, u)$.

The relation between the interior set systems and convex geometries was studied in [2]. However, the corresponding statement in [2] is not quite exact because it does not include the requirement that $\phi_u(H)$ is well-defined.

## References


[1] B. Mirkin and I. Muchnik, Layered clusters of tightness set functions, submitted (2001).

[2] I. Muchnik and L. Shvartser, Maximization of generalized characteristics of functions of monotone systems, *Automation and Remote Control*, **51**, 1562-1572 (1990).

[3] P.H. Edelman and R.E. Jamison, The theory of convex geometries, *Geom. Dedicata*, **19**, 247-270 (1985).

[4] H. Chernoff, Rational selection of decision functions, *Economica*, **22**, 422-443 (1954).

[5] P.H. Edelman and M.E. Saks, Combinatorial representation and convex dimension of convex geometries, *Order*, **5**, 23-32 (1988).

[6] B. Korte, L. Lovász, and R. Schraeder, *Greedoids*, Springer-Verlag, New York (1991).