Chapter 4: The Hypernetwork Model of Memory

4.1 Hypergraphs

유한 집합 $X = \{x_1, x_2, ..., x_n\}$에 대해서 $E = \{E_i \mid i \in I\}$를 ***X*의 부분집합 *E*들의 패밀리(family)**라고 정의한다. 패밀리 E가 다음조건을 만족할 때 ***E*는 *X* 상에 정의된 하이퍼그래프**라고 한다.

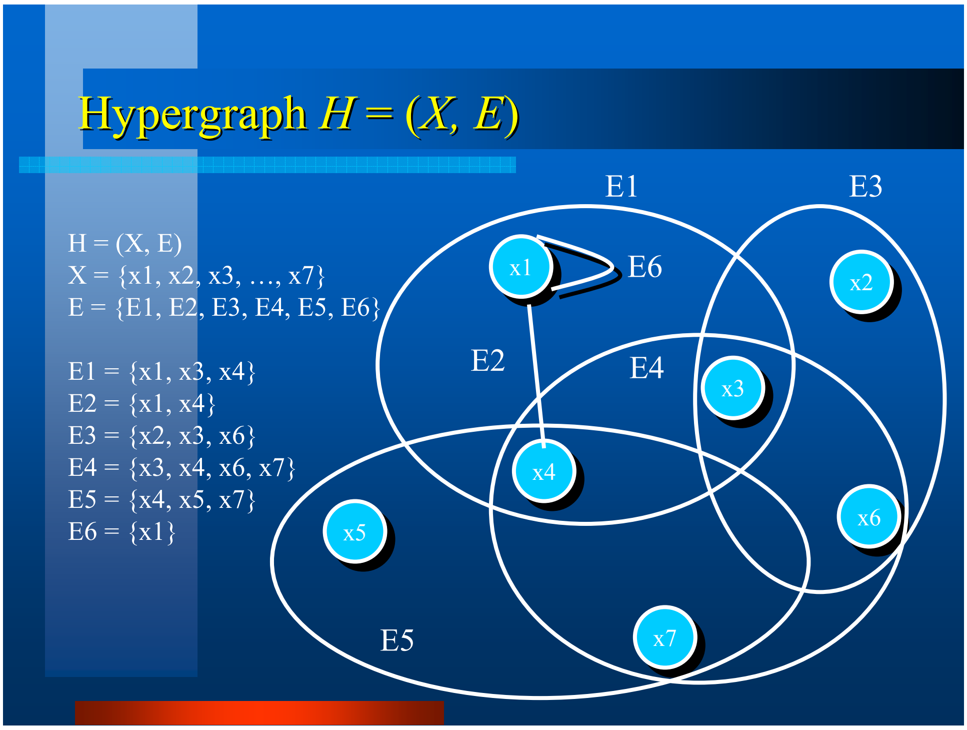
$$E_i \neq \varnothing,\ i \in I \qquad \bigcup_{i \in I} E_i = X \tag{1.1}$$

각각의 요소 $x_1, x_2, ..., x_n$을 정점(vertices)이라 하고 집합 $E_1, E_2, ..., E_m$을 하이퍼에지(hyperedge) 또는 에지(edges) 라 한다.

X의 멱집합(power set)을 $Pw(X) = \{E_i \mid E_i \subseteq X\} = 2^X$라 하자. 즉

$$Pw(X) = \{E_i \mid E_i \subseteq X\} = 2^X \tag{1.2}$$

하이퍼그래프 $H = (X, E)$는 정점의 집합 X와 에지의 집합 E으로 구성되어 있으며 여기서 E는 X의 멱집합 $Pw(X) = 2^X$의 한 부분집합이다. $|X| = n$을 이 **하이퍼그래프의 차수(order)**라 한다.



<그림> 7개의 정점과 6개의 하이퍼에지로 구성된 하이퍼그래프 $H=(X,E)=(\{x_1,x_2,...,x_7\},\{E_1,E_2,...,E_6\})$. 그림에서 $|E_i|>2$인 edge는 E_i에 속한 모든 정점들을 포함하는 원으로 표시하였다. $|E_i|=2$인 edge는 두 정점을 연결하는 선으로 표시하였으며 . $|E_i|=1$인 edge는 그래프상에서 loop으로 표시하였다.

Incidence Matrix for Hypergraph H

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
E_1	1		1	1			
E_2	1			1			
E_3		1	1			1	
E_4			1	1		1	1
E_5				1	1		1
E_6	1						

$X^{[m]}$을 정확히 m개의 원소를 가진 X의 부분집합이라 하자. 즉 $X^{[m]}$은 m-부분집합이다. ***m*-하이퍼그래프**는 정점의 집합 X와 에지의 집합 $E=X^{[m]}$로 구성되어 있다. 즉 $H^{[m]}=(X,X^{[m]})$이다.

하이퍼그래프의 에지가 모두 같은 수의 원소들로 구성되어 있으면 그 **하이퍼그래프는 균일(uniform)하다**고 한다. 하이퍼그래프 H의 모든 에지 $E_i\in E$의 원소수(cardinality)가 k일 때 H는 ***k*-uniform** 하다라고 한다.

정점 x_i의 차수 $d(x_i)$는 E에 속한 에지 중에 이 정점을 포함하고 있는 에지의 수로 정의된다. H의 모든 정점이 차수가 k일 때 H는 ***k*-regular 하이퍼그래프**라고 한다.

Note: 일반적인 그래프는 2-uniform 하이퍼그래프이다.

E_i가 모두 다른 하이퍼그래프를 **단순(simple) 하이퍼그래프**라 한다. 모든 i에 대해서 $|E_i|=2$이고 하이퍼그래프 H가 단순이면, H는 분리된 정점을 갖지 않는 단순 그래프(simple graph)이다.

하나의 하이퍼그래프에서 두 개의 정점을 포함하는 에지 E_i가 존재하면 이 **두 정점은 인접해 있다(adjacent)**고 한다. 두 개의 에지의 교집합이 공집합이 아니면 두 개의 에지는 인접해 있다(adjacent).

하이퍼그래프 **H = (X, E)의 Incidence Matrix**는 H의 에지를 나타내는 m개의 횡(row)과 H의 정점을 나타내는 n개의 열(column)로 구성된 매트릭스 $\left(a_j^i\right)$이다.

$$a_j^i=\begin{cases}1, & \text{if } x_j\in E_i,\\ 0, & \text{otherwise}\end{cases} \tag{1.3}$$

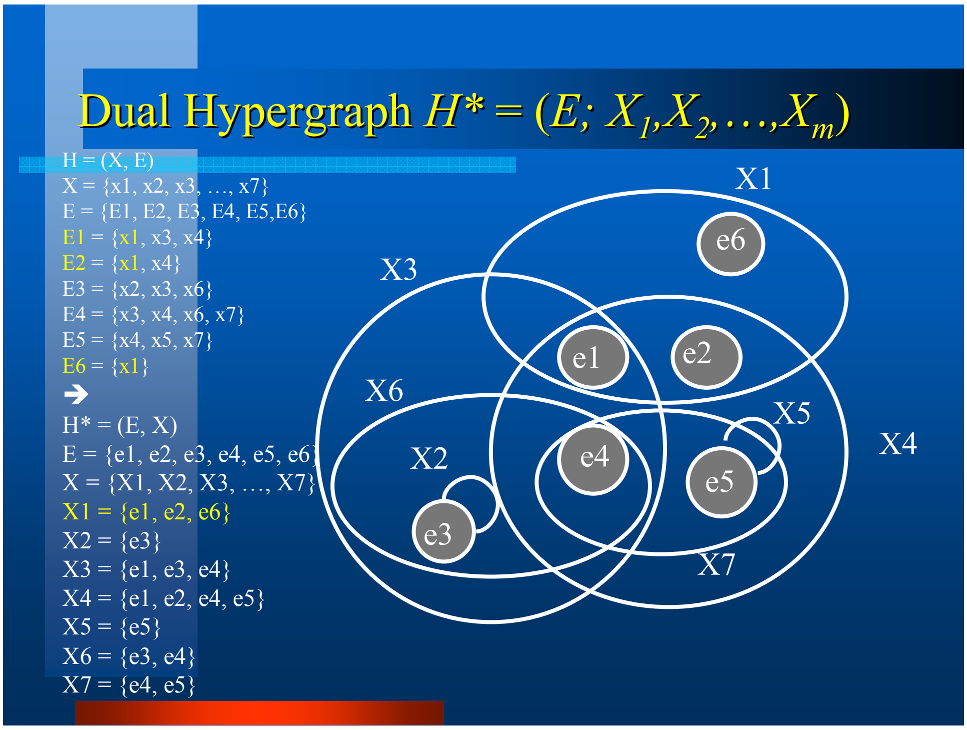
순수하게 0으로만 구성된 횡과 열이 존재하지 않은 (0-1) 매트릭스는 하이퍼그래프 H의

incidence matrix이다.

각 하이퍼그래프 $H = (X\,;E_1, E_2, ..., E_m)$에 대해서 다음과 같은 대응되는 하이퍼그래프 $H^* = (E\,;X_1, X_2, ..., X_m)$ 이 존재하며 H^*의 정점은 $e_1, e_2, ..., e_m$ 의 점들이고(각각 $E_1, E_2, ..., E_m$에 대응되는) 에지는 집합 $X_1, X_2, ..., X_n$ (각각 $x_1, x_2, ..., x_n$에 대응)이며 모든 j에 대해서

$$X_j = \{e_i \mid i \le m,\ E_i \ni x_j\} \tag{1.4}$$

이다. 따라서 $X_j \ne \varnothing$이고 $\bigcup_j X_j = E$이며 H^*는 하이퍼그래프이다.



<그림> 듀얼 하이퍼그래프 $H^* = (E, X)$.

Incedence Matrix for Dual Hypergraph H^*

	e_1	e_2	e_3	e_4	e_5	e_6
X_1	1	1				1
X_2			1			
X_3	1		1	1		
X_4	1	1		1	1	
X_5					1	
X_6			1	1		

X_7				1	1	

H^*는 H의 **듀얼 하이퍼그래프(dual hypergraph)**라 한다. 하이퍼그래프 H^*의 incidence matrix는 하이퍼그래프 H의 incidence matrix $\left(a_j^i\right)$의 transpose이다. 즉 $(H^*)^* = H$. H에 있는 두 개의 정점 x_j와 x_k가 인접하면 H^*에 있는 대응하는 에지 X_j와 X_k가 인접하다. H에 있는 두 개의 에지 E_i와 E_j가 인접하면 H^*에 있는 두 개의 정점 e_i와 e_j가 인접하다.

하이퍼그래프 H에서 공집합이 아닌 $S \subset X$에 대해 $r(S) = \max_i |S \cap E_i|$인 양수를 ***S*의 랭크(rank)**라 정의한다. 하이퍼그래프 H의 랭크는 $r(X)$로 정의된다. 각 i에 대해서 $E_i = r(X)$이면 H는 랭크 $r(X)$인 **균일(uniform) 하이퍼그래프**라고 한다. 따라서 분리된 정점을 갖지 않는 단순 그래프는 랭크 2인 균일 하이퍼그래프이다. 하이퍼그래프 H = (X, E)에 대해 패밀리 $F \subset E$에 의해서 생성된 하이퍼그래프 **(X_F, F)를 H의 일부(partial) 하이퍼그래프**라고 한다. 여기서 X_F는 다음과 같이 정의된다.

$${}_F = \bigcup_{E_i \in F} E_i \tag{1.5}$$

하이퍼그래프 H = (X, E)에 대해 집합 $A \subset X$에 의해서 생성된 하이퍼그래프 H_A = (A, E_A)를 H의 **부분 하이퍼그래프(subhypergraph)**라고 한다. 여기서 $E_A = \{E_i \cap A \mid E_i \in E;\ E_i \cap A \neq \varnothing\}$이다.

정수 $k > 0$에 대해 하이퍼그래프 $H = (X, E)$의 ***k*-섹션 (k-section)은** $H_{(k)} = (X, E_{(k)})$로 정의된다. 여기서 $E_{(k)}$는 다음과 같다.

$$E_{(k)} = \{F \mid F \subset X;\ 1 \leq |F| \leq k;\ F \subset E_i \text{ for some } E_i \in E\} \tag{1.6}$$

주어진 하이퍼그래프 $H = (X, E)$에 대해서 **정점 *x*의 차수(degree)**는

$$E_i \cap E_j = \{x\} \qquad (i, j \in J;\ i \neq j) \tag{1.7}$$

인 일부 패밀리(partial family) $(E_j \mid j \in J)$를 구성하고 $\{x\}$와 다른 최대 에지수로 정의된다. ***x*의 차수를 $d_{H(x)}$로 나타**낸다. $d_{H(x)} = 0$이면 x를 포함하는 유일한 에지는 $\{x\}$이며 그 역도 성립한다.

하이퍼그래프 $H = (X, E)$에서 **길이 *q*인 사슬(chain)**은 sequence (x_1, E_1, x_2, E_2, ..., E_q, x_{q+1})로 정의된다. 여기서

(1) $x_1, x_2, ..., x_q$은 모두 H의 서로 다른 정점이고

(2) $E_1, E_2, ..., E_q$은 모두 H의 서로 다른 에지이고

(3) $k = 1, 2, ..., q$에 대해 $x_k, x_{k+1} \in E_k$이다.

만약 $q < 1$이고 $x_{q+1} = x_1$이면 이 사슬은 길이 q인 사이클이라고 한다. 하이퍼그래프에서 정점 a로 시작해서 정점 b로 끝나는 사슬이 존재하면 이를 a++b로 표시한다. 또한 릴레이션 a++b는 equivalence relation이며 그 클래스는 하이퍼그래프의 **Connected Components**라고 한다. 만약 C_1이 에지 E와 교차하는 하나의 connected component이면 C_1은 E를 포함한다.

주어진 하이퍼그래프 $H = (X, E)$에 대해서 집합 $S \subset X$가 $|E_i| > 1$인 에지 E_i를 하나도 포함하지 않으면 ***S*는 안정적(stable)**이라고 한다. ***H*의 안정도(stability number)**는 H의 안정적인 집합의 최대 크기(maximum cardinality)로 정의된다. 하이퍼그래프 ***H*의 Chromatic Number** $\chi(X)$는 $|E_i| > 1$인 어떤 에지 E_i 도 모든 인접 정점과 같은색을 갖지 않도록 H의 정점들을 색칠하는데 필요한 최소 색깔수(smallest number of colors)로 정의된다. q-coloring 문제는 X를 각각 다른 색깔을 사용하는 q개의 안정된 집합 $S_1, S_2, ..., S_q$로 조각내는(partition) 문제이다. q-coloring이 가능한 하이퍼그래프를 **q-colorable**하다고 한다.

Example: Movie Script Data

There's nothing to tell

He's just some guy I work with

You're going out with the guy

...

➔

E1: There's nothing to

E2: nothing to tell

E3: He's just some

E4: just some guy

E5: some guy I

E6: guy I work

E7: guy I work with

E8: You're going out

E9: going out with

E10: out with the

E11: with the guy

...

➔

Incidence Matrix for the 3-Hypergraph

	x_1 There's	x_2 nothing	x_3 to	x_4 tell	x_5 He's	x_6 just	x_7 some	x_8 guy	x_9 I	x_{10} work	x_{11} with	x_{12} You're	...
E_1	1	1	1										
E_2		1	1	1									
E_3					1	1	1						
E_4						1	1	1					

E_5							1	1	1				
E_6								1	1	1			
E_7									1	1	1		
E_8												1	1
E_9													
...													

Incidence Matrix for the Dual 3-Hypergraph

	e_1 There's nothing to	e_2 nothing to tell	e_3 He's just some	e_4 just some guy	e_5 some guy I	e_6 guy I work	e_7	e_8	e_9	e_{10}	...
X_1 There's	1										
X_2 nothing	1	1									
X_3 to	1	1									
X_4 tell		1									
X_5 He's			1								
X_6 just			1	1							
X_7 some			1	1	1						
X_8 guy				1	1	1					
X_9 I					1	1					
X_{10} work						1					
...											

H

노드수 = 단어수 $|V| = 30000$

에지수 = 문장수 * (문장길이 − (k-1)))

H^*

노드수 = 문장수 * (문장길이 − (k-1)))

에지수 = 단어수 $|V| = 30000$