

MkCP (Maximum k -Club Problem)를 위한 휴리스틱 기반 알고리즘

김소정¹, 김찬수², 한근희^{2*}

¹공주대학교 수학과 석사과정, ²공주대학교 응용수학과 교수

A Heuristic-Based Algorithm for Maximum k -Club Problem

SoJeong Kim¹, ChanSoo Kim², KeunHee Han^{2*}

¹Master student, Department of Applied Mathematics, Kongju National University

²Professor, Department of Applied Mathematics, Kongju National University

요약 k -club은 소셜 네트워크 분석에서 다양한 형태의 소셜 그룹을 설명하기 위해 제안된 그래프 모델 중 하나로, 단순 그래프에서 부분 정점 집합 S 에 의한 유도 부분그래프(Induced subgraph)의 지름이 k 보다 작거나 같은 경우 S 를 k -club이라 한다. 본 논문에서는 유전알고리즘을 이용하여 그래프에서 크기가 최대인 k -club을 찾는 문제인 MkCP(Maximum k -Club Problem)을 계산하는 HGA+DROP 알고리즘을 제안한다. 본 알고리즘은 k -club을 위한 휴리스틱 알고리즘 k -CLIQUE & DROP을 변형하고 휴리스틱 유전 알고리즘(HGA)을 사용해 한 번의 수행으로 복수 개의 k -club을 구하였다. 기존 알고리즘의 결과와 비교하기 위해 DIMACS 그래프들에 대하여 k 가 2, 3, 4 그리고 5일 때 MkCP를 계산하였다.

주제어 : k -club, MkCP(Maximum k -club problem), 휴리스틱 알고리즘, 유전 알고리즘, DIMACS

Abstract Given an undirected simple graph, k -club is one of the proposed structures to model social groups that exist in various types in Social Network Analysis (SNA). Maximum k -Club Problem (MkCP) is to find a k -club of maximum cardinality in a graph. This paper introduces a Genetic Algorithm called HGA+DROP which can be used to approximate maximum k -club in graphs. Our algorithm modifies the existing k -CLIQUE & DROP algorithm and utilizes Heuristic Genetic Algorithms (HGA) to obtain multiple k -clubs. We experiment on DIMACS graphs for $k = 2, 3, 4$ and 5 to compare the performance of the proposed algorithm with existing algorithms.

Key Words : k -club, MkCP, Heuristic algorithm, Heuristic genetic algorithm, DIMACS

*This work was supported by Building Data for AI Learning(Video Narrative and Q&A Data) from National Information Society Agency(NIA).

*Corresponding Author : Keunhee Han(kehan@kongju.ac.kr)

Received August 4, 2021

Accepted October 20, 2021

Revised September 16, 2021

Published October 28, 2021

1. 서론

단순 그래프 $G = (V, E)$ 는 정점 집합 $V = \{1, \dots, n\}$ 및 간선집합 $E \subseteq \{(u, v) \mid u, v \in V\}$ 으로 구성된다. 그래프 G 에 대하여 $\deg_G(u)$ 는 정점 u 의 차수를, $d_G(u, v)$ 는 G 내 두 개 정점 u, v 사이의 최단거리를 의미하고, $\text{diam}(G) = \max\{d_G(u, v) \mid u, v \in V\}$ 는 G 의 지름(diameter)을 나타낸다. 그래프 $G = (V, E)$ 및 양의 정수 k 에 대하여 $G^k = (V, E^k)$ 는 정점 집합 V 및 간선 집합 $E^k = \{(u, v) \mid d_G(u, v) \leq k\}$ 로 구성되며, 정점 집합 $S \subseteq V$ 에 대하여 $G[S] = (S, E \cap (S \times S))$ 는 G 의 유도부분그래프(Induced subgraph)를 나타낸다.

그래프 $G = (V, E)$ 에 대하여 정점 집합 $S \subseteq V$ 및 k 를 양의 정수라 하자. 만일 $G[S]$ 가 완전그래프라면 S 를 G 의 clique라 하고 $u, v \in S$ 에 대하여 $d_G(u, v) \leq k$ 이면 S 를 G 의 k -clique라 하며 $\text{diam}(G[S]) \leq k$ 이면 S 를 G 의 k -club이라 한다. MCP (Maximum Clique Problem)는 그래프 내 크기가 최대인 clique를 찾는 문제인 반면에, k MCP (Maximum k -Club Problem)는 그래프에서 크기가 최대인 k -club을 찾는 문제로 [1]에 따르면 $k \geq 2$ 인 경우 k MCP는 NP-hard 계열에 속하는 문제이다. 정리 1은 [2]에서 제시된 k -club과 k -clique의 관계를 나타낸 것으로 k -club은 또한 k -clique임을 보이지만 일반적으로 역은 성립하지 않는다. 예를 들어, Fig 1에서 $\{1, 2, 3, 4, 5\}$ 는 2-clique이지만 2-club은 아니며, $\{1, 2, 3, 4\}$ 는 2-club 및 2-clique임을 알 수 있다.

정리 1. $S \subseteq V$ 가 $G = (V, E)$ 의 k -club 이라면 S 는 또한 G 의 k -clique 이다. [2]

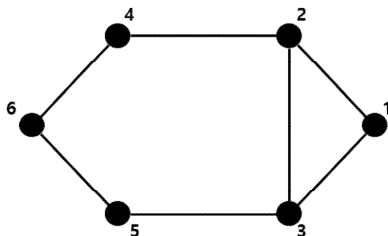


Fig. 1. graph $G = (V, E)$

k -club은 유전성(heredity property)을 만족하지 않

기 때문에 부분 집합에 대해 닫혀있지 않아 k -club의 부분 집합이 k -club 임은 보장할 수 없으며, 본 성질로 인해 극대(maximal) k -club임을 검사하는 문제 또한 NP-complete 계열에 속한다[3,4].

소셜 네트워크 분석(SNA)에서 응집 부분집단(cohesive subgroup)이라 불리는 밀집 구조(dense structure)를 식별하는 것은 본 분야의 주된 관심사 중 하나로, 이를 통해 복잡한 네트워크에 대한 통찰력을 얻을 수 있으며, 이는 SNA 이외에도 네트워크 분석을 수행하는 다양한 분야에서 응집 부분집단을 적용하는 요인이 되었다[2,5-8]. 응집 부분 집단을 반영하는 그래프 모델로 초기에는 clique를 활용하였지만 clique의 형태 및 정의가 응용 분야에서 매우 드물게 나타나는 구조로 효용성이 떨어져 clique 보다 엄격하지 않은 정의를 가지는 그래프 모델의 필요성이 제기되었다[2,9]. 이에 따라 클릭 완화(clique relaxation)라 불리는 다양한 그래프 모델이 제안되었고, 본 논문에서는 클릭 완화 모델 중 하나인 k -club과 k MCP에 대한 연구를 수행하였다.

클릭 완화에 속하는 그래프 모델을 사용한 연구가 다양한 분야에서 수행되어졌다. [10] 및 [11]은 각각 데이터 마이닝과 생물 네트워크에서 클릭 완화 모델을 사용해 그래프 클러스터링(clustering) 문제를 분석하였다. NP-hard 문제인 k MCP를 위해 휴리스틱, 선형 계획법 그리고 Exact Algorithm 개발 외에도 여러 측면으로 연구가 수행되어졌다. [12]는 3가지 휴리스틱 알고리즘인 CONSTELLATION, DROP 및 k -CLIQUE & DROP을 제시하였으며, 이 중 DROP 알고리즘은 이후 많은 연구에 영향을 주었다. [13]는 두 단계로 구성된 휴리스틱 알고리즘인 constructive 및 elimination heuristic 등을 제안하였다. [3]는 k -club의 극대성(maximality) 검사를 위한 충분조건과 k MCP를 위한 VNS 알고리즘을 제시하였다. VNS 알고리즘은 초기 해를 구하기 위해 [12]의 휴리스틱을 포함한 3가지 알고리즘들을 응용하였다. [14]는 그래프의 정점 개수가 n 일 때 $O(1.62^n)$ 시간 내에 k MCP를 계산하는 알고리즘을 제안하고 새로운 휴리스틱 알고리즘인 IDROP을 소개하였다. [1]은 앞서 언급한 것과 같이 k MCP가 NP-hard에 속하는 문제임을 보이고, 정수 선형 계획법(Integer Linear Programming)을 사용해 k MCP를 수식으로 변형하였다. 이후 선형 계획법을 사용해 k MCP를 다양한 수식으로 변형하는 연구가 수행되어졌다 [15-18]. [19,20]는 k 가 2인 경우에 한하여 k MCP를 위한 연구를 수행하였다. [19]는 그래프에

존재하는 가장 큰 2-club만 찾는 것이 아닌 모든 2-club을 찾는 연구를 수행하였고 [20]은 다양한 조건 하에서 2-club 문제가 NP-hard에 속하는 문제임을 보였으며, distance to cograph에 대한 고정 파라미터 알고리즘(fixed-parameter algorithm)을 제시하고 이에 대한 성능을 distance to cluster graphs에 대한 알고리즘과 비교하였다.

MkCP 계산을 위해 여러 방향으로 연구가 수행되었으나 유전성이 없는 문제의 특성 때문에 k 를 특정 정수 값으로 한정하거나 수식으로 변형해 해결하는 연구가 활발히 이뤄졌다. 본 논문에서는 [12]에서 제안한 휴리스틱 알고리즘인 k -CLIQUE & DROP에 다음과 같이 두 가지의 변형을 주어 더욱 효율적으로 MkCP를 계산하는 방법을 제시한다. 첫째, 그래프 G 의 k -clique가 아닌 G^k 의 clique를 찾도록 변형하며 이를 위해 휴리스틱 유전 알고리즘(Heuristic Genetic Algorithm, HGA)[21]을 적용하였고, 둘째, 유전 알고리즘과 달리 한 번의 HGA 실행으로 다수의 clique를 얻을 수 있도록 수정해 알고리즘의 효율성을 높였다.

유전 알고리즘은 다윈의 자연 선택설을 기반으로 고안된 알고리즘으로 최적화 및 탐색 문제 해결에 사용하고, 탐색 공간의 제약에서 비교적 자유롭기 때문에 복잡한 문제 해결에 유용하여 다양한 분야에서 활용되고 있다[22]. 유전 알고리즘은 염색체(chromosome)의 집합인 해집단(population)과 유전 연산자들을 사용하여 탐색 공간에서 최적값을 찾는다. 염색체는 문제의 후보 해를 알고리즘에 맞게 변형한 것으로 이들은 각각 적합도(fitness)를 가진다. 적합도는 목표 함수에 대한 값으로 염색체가 문제에 얼마나 적합한 지를 나타내는 지표이다. 유전 연산자는 선택(Selection), 교배(Crossover) 및 돌연변이(Mutation) 등이며 현세대의 해집단으로부터 다음 세대의 해집단을 생성하는 기능들을 담당한다. 선택 연산은 해집단에서 다음 세대를 생성할 부모 염색체들 그들의 적합도에 따라 선택하고, 교배 연산을 통해 부모 염색체들로부터 새로운 자식 염색체들을 생성하며, 돌연변이 연산은 매우 드문 확률로 염색체의 유전자를 무작위로 변형시킨다. 교배 연산을 통해 우수한 부모의 유전자가 자손에게 전달되어 더욱 우수한 염색체가 생성되는 것을 목표로 하며, 매우 작은 확률로 적용되는 돌연변이 연산은 염색체의 유전자를 변형시켜 지역 최적해로부터 벗어날 수 있는 기회를 제공한다.

본 논문의 나머지 부분은 다음과 같이 구성된다. 2절에서는 MkCP를 위한 효율적인 알고리즘을 제시하고 3

절에서는 제시된 알고리즘의 성능을 다른 연구의 결과들과 비교 분석하였다. 끝으로 4절에서 결론을 제시한다.

2. HGA+DROP 알고리즘

본 절에서는 [12]의 휴리스틱 알고리즘인 k -CLIQUE & DROP을 기반으로 MkCP를 계산하는 새로운 HGA+DROP 이라는 알고리즘을 제시한다. k -CLIQUE & DROP 알고리즘은 2 단계로 구성된 휴리스틱 알고리즘으로, 그래프 G 로부터 (1) 크기가 큰 k -clique C 를 계산한 후 (2) C 내에서 DROP 알고리즘을 통해 k -club을 계산하며, 이는 정리 1이 제시하는 k -clique 및 k -club 사이의 관계를 이용한 것이다.

HGA+DROP은 k -CLIQUE & DROP의 첫 번째 단계를 변형한 형태로 구성되었다. 그래프 G 의 k -clique인 정점 집합이 그래프 G^k 에서 clique인 성질을 이용하여, G 의 크기가 최대인 k -clique를 찾는 대신에 G^k 의 크기가 최대인 clique를 찾는 문제 즉, G^k 의 MCP 문제로 변환하고 이를 위해 [21]에서 제안한 휴리스틱 유전 알고리즘(HGA)이 적용된다. HGA를 통해 찾은 clique를 사용하여 유도 부분그래프를 생성하고 이에 DROP을 적용해 MkCP를 계산한다.

G^k 는 그래프 G 의 모든 정점 사이의 최단거리를 계산하여 얻을 수 있는 그래프로, 실험을 통해 HGA+DROP의 수행시간을 증가시키는 주요인이 G^k 계산이라는 것을 확인하였고, 이 과정을 병렬 연산을 사용해 실행함으로써 계산 시간을 감소시켰다.

HGA는 휴리스틱 유전 알고리즘으로, 적합도 평가 전에 지역 최적해 연산을 수행함으로써 무작위로 생성된 염색체(chromosome)는 항상 clique임을 보장한다. 염색체는 0-1 벡터 v 로 구성되며 $v[i] = 1$ 이면 부분 정점 집합에 포함된 정점임을 나타내고, 0이면 포함되지 않은 정점임을 나타낸다. 각 염색체의 적합도(fitness)는 clique인 부분 정점 집합의 크기로 결정된다. [21]에서 제안된 휴리스틱 연산은 Relax, Repair 및 Extend 등 세 단계로 구성된 알고리즘으로써 모든 염색체에 각각 적용된다. Relax는 그래프의 모든 정점 중 부분 그래프에 포함되지 않은 정점을 무작위로 부분 그래프에 추가하는 단계로 이 과정을 통해 부분 정점 집합의 크기를 키운다. Repair는 부분 그래프의 정점 집합을 clique로 구성하기 위하여 집합 내 정점을 제거하는 단계이고,

Extend는 Repair를 통해 clique가 된 부분 정점 집합의 크기를 키우는 단계로 부분 정점 집합에 포함되지 않은 정점 중 집합 내 모든 정점과 인접한 정점을 집합에 추가하는 방법을 사용한다. Repair와 Extend는 처음 제거하거나 추가하는 시작 정점에 따라 결과가 달라지기 때문에 무작위로 시작 정점을 선택하여 더욱 다양한 clique가 생성될 수 있게 한다.

본 알고리즘에서 HGA는 단순히 MCP를 해결하기 위함이 아니라 최대 clique를 구하고 이를 통해 MKCP를 계산하기 위해 사용하기 때문에, 우리는 다수의 clique를 한 번의 HGA 수행으로 구함으로써 알고리즘의 효율성을 높였다. 따라서 HGA는 clique 집합을 결과로 돌려주며, 이 때 clique 집합은 크기는 모두 동일하지만 각 정점 집합의 요소가 다른 clique로 구성된다.

Algorithm 1. DROP

INPUT : $G = (V, E)$, positive integer k

OUTPUT : k -club G'

```

1 while(true)
2   compute  $d_G(u,v)$  for  $u, v \in V(G)$ 
3   for  $v = 1$  to  $|V(G)|$  do
4      $q_v = \{u \mid d_G(u,v) > k\}$ 
5   end for
6   Exit while loop if  $q_i = 0$  for all  $i \in V(G)$ 
7    $W = \{u \mid q_u > 0\}$ 
8   Let  $v$  be the vertex of  $W$  such that  $\deg_G(v)$  is minimum.
   (Ties are broken arbitrary)
9    $G = G - v$ 
10 end while

```

Fig. 2. DROP algorithm

k -CLIQUE & DROP의 두 번째 단계인 DROP 알고리즘 또한 [12]이 제안한 휴리스틱 중 하나로 그래프의 정점 중 조건에 따라 정점 하나를 선택하고, 그래프에서 선택된 정점과 이를 끝점으로 가지는 간선을 모두 제거하여 그래프가 k -club이 될 때까지 반복하는 알고리즘이다. 본 논문에서는 HGA로 구한 clique 집합을 통해 생성한 유도 부분그래프에 DROP을 적용하여 k -club을 구하며 Fig 2에 DROP 알고리즘 과정을 나타내었다.

Fig 3은 HGA+DROP 알고리즘을 보여준다. 단계 1-2은 G 내 최단거리를 계산하고 G^k 를 계산하는 과정이고, 단계 3-13은 HGA의 수행 과정을 나타낸다. 단계 5의 CS 는 clique 집합을 의미하며, evaluation 함수는 해집합 내 모든 개체의 적합도를 계산하고 현재까지 찾은 가장 좋은 적합도와 비교하여 더 좋은 적합도를 찾았다면 CS 를 갱신하고 그렇지 않다면 현재 가장 좋은 적합도와 동일한 적합도를 갖

는 개체를 CS 에 저장한다. 마지막으로 단계 14-17은 HGA를 통해 찾은 clique들을 사용해 DROP을 적용하는 과정을 보여준다.

Algorithm 2. HGA+DROP

INPUT : $G = (V, E)$, positive integer k

```

1 compute  $d_G(u,v)$  for  $u, v \in V(G)$ 
2 Let  $G^k = (V, E^k)$  where  $E^k = \{(u, v) \mid d_G(u,v) \leq k\}$ 
3  $g = 0$ 
4 Initial population  $P_g$ 
5  $CS = \text{evaluation}(P_g)$ 
6 while( termination condition not met )
7    $g = g + 1$ 
8    $parent_{g-1} = \text{selection}(P_{g-1})$ 
9    $P_g = \text{crossover}(parent_{g-1})$ 
10  mutation( $P_g$ )
11  heuristic( $P_g$ )
12   $CS = \text{evaluation}(P_g)$ 
13 end while
14 for each clique  $c$  in  $CS$  do
15   Let  $G[c]$  be a subgraph of  $G$  induced by  $c$ 
16   DROP( $G[c]$ ,  $k$ )
17 end for

```

Fig. 3. HGA+DROP algorithm

3. 실험 및 결과 비교

제시된 알고리즘은 Java 언어를 이용하였으며 총 3개의 클래스(Find_club, Graph_k 및 HGA)를 사용해 구현하였다. Find_club은 다른 두 개 클래스를 호출하여 사용하는 메인 클래스이며 Graph_k는 G^k 는 Java의 Thread를 사용해 계산하는 병렬 프로그램 클래스이고 마지막으로 HGA는 주어진 G^k 에서 clique를 구하는 휴리스틱 유전 알고리즘을 구현한 클래스이다.

Table 1. parameters of HGA

Parameter	
max generation	50
population size	500
tournament size	17
crossover rate	0.9
mutation rate	0.1

실험 환경은 Intel 3.50 GHz 프로세서와 16 GB 램이 설치된 데스크탑 PC에서 실행하였으며 실험은 10번째 DIMACS Implementation challenge [23]에서 제시된 그래프들 중 20개를 사용하여 HGA+DROP 알고리즘의

Table 2. size of k -clubs for $k = 2$ using VNS, B&B and HGA+DROP algorithms

Graph	V	E	VNS	B&B	HGA+DROP	
			2-club size	2-club size	2-club size	CPU (s)
karate	34	78	18	18	18	0.617
dolphins	62	159	13	13	13	0.462
polbooks	105	441	28	28	28	0.646
adjnoun	112	425	50	50	50	0.809
football	115	613	16	16	16	0.648
jazz	198	2742	103	103	103	3.324
celegans-metabolic	453	2025	238	238	238	8.536
email	1133	5451	72	72	72	18.104
polblogs	1490	16715	352	352	352	47.453
netscience	1589	2742	35	35	35	32.304
add20	2395	7462	124	124	124	100.301
data	2851	15093	18	18	18	137.397
uk	4824	6837	5	4	5	395.489
power	4941	6594	20	20	20	317.237
add32	4960	9462	32	32	32	419.824
hep-th	8361	15751	51	51	51	934.584
whitaker3	9800	28989	9	9	9	1711.662
crack	10240	30380	10	10	10	1885.510
PGPgiantcompo	10680	24316	206	-	206	2070.424
cs4	22499	43858	5	-	6	10046.628

Table 3. size of k -clubs for $k = 3$ using VNS, B&B and HGA+DROP algorithms

Graph	V	E	VNS	B&B	HGA+DROP	
			3-club size	3-club size	3-club size	CPU (s)
karate	34	78	25	25	25	0.649
dolphins	62	159	29	29	29	0.575
polbooks	105	441	53	53	53	0.852
adjnoun	112	425	82	82	82	1.801
football	115	613	58	58	56	13.300
jazz	198	2742	174	174	174	5.543
celegans-metabolic	453	2025	371	371	371	20.600
email	1133	5451	215	201	198	130.879
polblogs	1490	16715	352	-	774	1062.557
netscience	1589	2742	54	54	54	39.411
add20	2395	7462	671	671	671	164.105
data	2851	15093	32	28	32	138.101
uk	4824	6837	8	8	8	394.174
power	4941	6594	30	30	30	406.392
add32	4960	9462	96	99	99	424.936
hep-th	8361	15751	120	120	120	1146.315
whitaker3	9800	28989	15	13	15	1706.994
crack	10240	30380	17	15	17	1880.984
PGPgiantcompo	10680	24316	273	-	422	2126.167
cs4	22499	43858	9	-	10	9947.619

성능을 VNS (Variable Neighborhood Search) 알고리즘[3]과 B&B (Branch and Bound) 알고리즘[4]의 결과와 비교 분석하였다.

HGA의 선택 연산자는 토너먼트 선택 (Tournament Selection), 교배 연산자는 균일 교배 (Uniform Crossover) 그리고 돌연변이 연산자는 비트 플립 돌연변이 (Bit-Flip Mutation)를 사용하였다. HGA의 매개 변수인 종료 조건으로 최대 세대 수 (max generation)를 50으로 설정하였고 Table 1 은 실험에 적용된 휴리스틱 유전 알고리즘의 파라미터들을 보여준다.

Table 2 및 3은 DIMACS 그래프에 대한 VNS 알고리즘, B&B 알고리즘의 k -club 크기 및 HGA+DROP의 k -club 크기와 수행시간을 보여준다. 2-club의 경우 HGA+DROP의 결과가 이전 연구의 결과들과 동일하고, 그래프 cs4에 대해 기존 결과보다 큰 2-club 크기를 구하였다. k 가 3일 때 몇 개의 그래프에 대해 VNS의 결과보다 좋지 않은 값을 얻었으나 polblogs, PGPgiant compo 및 cs4에서 기존의 결과들과 비교하여 매우 우월한 결과를 얻었으며 그 외 그래프에 대해서는 동일한 결과를 계산하였다. 비교적 정점 수가 적은 그래프에 대해서 VNS와 B&B 알고리즘에 비해 작은 크기의 k -club들이 계산되었지만 polblogs, PGPgiantcompo 및 cs4

와 같이 정점이 많고 밀집도가 낮은 그래프에 대해 매우 좋은 결과를 얻었기 때문에 HGA+DROP이 k CP 계산 을 우수하게 하고 있다고 판단된다. 전체적으로 VNS와 B&B 알고리즘의 결과와 비교해 대부분의 그래프에서 동일한 결과를 얻을 수 있었으며 정점 수가 많은 그래프에서 우수한 성능을 보였다. 또한 HGA+DROP의 수행 시간을 보았을 때 대부분의 그래프들에 대해 30분 내에 k CP를 계산하였음을 알 수 있다.

Table 4 는 k 가 4, 5일 때 HGA+DROP의 결과 및 수행시간을 나타내며, 이러한 k 값에 대한 연구는 조사한 바에 따르면 본 논문에서 처음 제시하는 결과이다. 몇 그래프의 k CP 값은 그래프의 정점 개수와 동일하며 이는 그래프의 지름이 k 와 동일하거나 작아서 발생하는 현상으로 그래프 전체가 k -club 임을 의미한다. 대부분의 그래프가 유사한 시간 내에 4-club과 5-club를 계산 하였으나 5개 그래프 email, polblogs, add20, hep-th 및 PGPgiantcompo에서 4-club과 5-club을 구할 때 수행 시간 차이가 많이 발생하는 것을 알 수 있으며, 이 중 polblogs는 k 가 4일 때 더 오랜 시간이 소요되었다. polblogs를 제외한 4개 그래프는 k 가 5일 때 k -club을 구하는 시간이 k 가 4인 경우보다 약 2배 이상 증가하였다. 이들의 경우 5-club의 크기가 원 그래프의 정점 수

Table 4. size of k -clubs for $k = 4$ and 5 using HGA+DROP algorithm

Graph	V	E	$k = 4$		$k = 5$	
			club size	CPU (s)	club size	CPU (s)
karate	34	78	33	0.747	34	0.789
dolphins	62	159	40	0.633	50	0.697
polbooks	105	441	68	1.107	95	1.405
adjnoun	112	425	107	1.715	112	1.737
football	115	613	115	1.896	115	1.762
jazz	198	2742	192	5.263	197	5.159
celegans-metabolic	453	2025	432	38.041	445	28.408
email	1133	5451	643	837.067	1003	2204.080
polblogs	1490	16715	1127	2513.185	1211	1133.219
netscience	1589	2742	85	36.685	116	39.750
add20	2395	7462	1454	630.570	2067	1702.593
data	2851	15093	52	116.109	73	123.838
uk	4824	6837	14	332.398	21	352.522
power	4941	6594	61	352.635	94	372.845
add32	4960	9462	268	365.436	499	398.193
hep-th	8361	15751	344	1073.007	764	2197.803
whitaker3	9800	28989	23	1456.258	33	1534.865
crack	10240	30380	31	1617.493	47	1692.797
PGPgiantcompo	10680	24316	1161	2199.572	1988	11862.473
cs4	22499	43858	18	8825.101	27	9120.815

와 큰 차이가 없거나 4-club과의 크기 차이가 300이상 발생하여 DROP 알고리즘을 수행하는데 오랜 시간이 소요되어 이로 인해 전체 수행시간이 상당히 증가하였으며 polblogs의 경우 k 가 4일 때 수행 시간이 오래 걸린 이유는 HGA를 통해 계산한 clique의 개수가 5-club을 구할 때 보다 많아 발생한 현상이다.

4. 결론

k -club은 그래프의 밀집된 구조를 나타내는 클리크 완화(clique relaxation) 그래프 모델 중 하나이고, MkCP는 그래프에 존재하는 가장 큰 k -club을 찾는 문제이다. MkCP는 NP-hard에 속하는 문제로 유전성(heredity)이 없는 특징으로 인해 이를 정확히 계산하는 알고리즘 개발에 어려움을 겪고 있다. 또한 일반적인 축소 규칙(reduction rule)이 적용될 수 없기 때문에 변형된 축소 규칙을 구하거나 적절한 경계값을 설정하려는 연구가 수행되어졌다.

본 논문에서는 휴리스틱 알고리즘인 k -CLIQUE & DROP를 바탕으로 MkCP를 계산하는 방법을 제안하였다. 이 때 그래프 G 에서 k -clique를 구하는 것이 아닌 G^k 에서 clique를 구하는 것으로 알고리즘을 변형하고, clique를 구하는 방법으로 휴리스틱 유전 알고리즘(HGA)을 사용하였으며 한 번의 실행으로 다수의 clique를 구함으로써 다양한 k -club을 얻을 수 있도록 하였다. HGA+DROP을 DIMACS 그래프에 적용해 k -club 크기와 수행 시간을 측정하였다. k 가 2일 때 기존 알고리즘과 비교하여 동일한 결과를 빠른 시간 내에 얻었으며, k 가 3일 때 기존 결과 보다 정점 수가 많고 밀도가 낮은 그래프에서 매우 우수한 성능으로 MkCP를 계산하였다. 4-club과 5-club을 비교하였을 때, HGA를 통해 다수의 clique를 구하는 것이 한 번의 수행으로 다양한 MkCP를 계산하는 효율성을 높이기도 하지만, clique를 구성하는 정점 수와 HGA를 사용해 찾은 clique의 수에 따라 전체 알고리즘의 수행 시간에 많은 영향을 미치는 것을 확인하였다.

MkCP 문제는 밀집도가 낮고 정점이 많은 그래프에서 더욱 어려운 문제로 알려져 있지만 이러한 특징을 가지는 DIMACS 그래프에서 HGA+DROP은 우수한 결과를 얻었다. 다만 비교적 난이도가 낮은 그래프에서 적절한 MkCP를 계산하지 못하는 현상에 대해 HGA의 매개변수 수정 또는 다른 알고리즘 고안하는 등의 연구가 필요

하다고 보여 진다.

REFERENCES

- [1] J. M. Bourjolly, G. Laporte & G. Pesant. (2002). An exact algorithm for the maximum k -club problem in an undirected graph. *European Journal of Operational Research*, 138(1), 21–28. [https://doi.org/10.1016/S0377-2217\(01\)00133-3](https://doi.org/10.1016/S0377-2217(01)00133-3)
- [2] R. J. Mokken. (1979). Cliques, clubs and clans. *Quality & Quantity*, 13(2), 161–173.
- [3] S. Shahram, & S. Butenko. (2013). Algorithms for the maximum k -club problem in graphs. *Journal of Combinatorial Optimization*, 26(3), 520–554. <https://doi.org/10.1007/s10878-012-9473-z>
- [4] F. M. Pajouh & B. Balasundaram. (2012). On inclusionwise maximal and maximum cardinality k -clubs in graphs. *Discrete Optimization*, 9(2), 84–97. <https://doi.org/10.1016/j.disopt.2012.02.002>
- [5] W. Stanley & K. Faust. (1994). *Social network analysis: Methods and applications*. Cambridge : Cambridge University Press
- [6] B. Balabhaskar, S. Butenko & S. Trukhanov. (2005). Novel approaches for analyzing biological networks. *Journal of Combinatorial Optimization*, 10(1), 23–39. <https://doi.org/10.1007/s10878-005-1857-x>
- [7] B. Balabhaskar, S. Butenko & I. V. Hicks. (2011). Clique relaxations in social network analysis: The maximum k -plex problem. *Operations Research*, 59(1), 133–142. <https://doi.org/10.1287/opre.1100.0851>
- [8] B. Vladimir, S. Butenko & os P. M. Pardalos. (2006). Mining market data: A network approach. *Computers & Operations Research*, 33(11), 3171–3184. <https://doi.org/10.1016/j.cor.2005.01.027>
- [9] R. D. Alba. (1973). A graph-theoretic definition of a sociometric clique. *Journal of Mathematical Sociology*, 3(1), 113–126.
- [10] B. Balasundaram & F. M. Pajouh. (2013). Graph theoretic clique relaxations and applications. *Handbook of combinatorial optimization*, 1559–1598.
- [11] B. Balasundaram, S. Butenko & S. Trukhanov. (2005). Novel approaches for analyzing biological networks. *Journal of Combinatorial Optimization*, 10(1), 23–39. <https://doi.org/10.1007/s10878-005-1857-x>
- [12] J. M. Bourjolly, G. Laporte & G. Pesant. (2000). Heuristics for finding k -clubs in an undirected graph. *Computers & Operations Research*, 27(6), 559–569. [https://doi.org/10.1016/S0305-0548\(99\)00047-7](https://doi.org/10.1016/S0305-0548(99)00047-7)
- [13] M. T. Almeida & F. D. Carvalho. (2014). Two-phase heuristics for the k -club problem. *Computers & operations research*, 52, 94–104.

<https://doi.org/10.1016/j.cor.2014.07.006>.

- [14] M. S. Chang, L.J. Hung, C. R. Lin & P. C. su. (2013). Finding large k -clubs in undirected graphs. *Computing*, 95(9), 739-758.
DOI : 10.1007/s00607-012-0263-3
- [15] M. T. Almeida & F. D. Carvalho. (2012). A comparison of integer models for the k -club problem. *Working Paper CIO 3/2012*, Universidade de Lisboa.
- [16] M. T. Almeida & F. D. Carvalho. (2014). An analytical comparison of the LP relaxations of integer models for the k -club problem. *European Journal of Operational Research*, 232(3), 489-498.
<https://doi.org/10.1016/j.ejor.2013.08.004>
- [17] A. Veremyev & V. Boginski. (2012). Identifying large robust network clusters via new compact formulations of maximum k -club problems. *European Journal of Operational Research*, 218(2), 316-326.
<https://doi.org/10.1016/j.ejor.2011.10.027>
- [18] M. T. Almeida & F. D. Carvalho. (2012). Integer models and upper bounds for the 3-club problem. *Networks*, 60(3), 155-166.
<https://doi.org/10.1002/net.21455>
- [19] S. Laan et al. (2012). *Fast finding of 2-clubs*. University of Amsterdam, Amsterdam.
- [20] S. Hartung, C. Komusiewicz, A. Nichterlein & O. Suchý. (2015). On structural parameterizations for the 2-club problem. *Discrete Applied Mathematics*, 185, 79-92.
<https://doi.org/10.1016/j.dam.2014.11.026>
- [21] E. Marchiori. (1998). A simple heuristic based genetic algorithm for the maximum clique problem. *In Symposium on Applied Computing: Proceedings of the 1998 ACM symposium on Applied Computing* (Vol. 27, pp. 366-373).
- [22] Z. Michalewicz. (2013). *Genetic algorithms+ data structures= evolution programs*. Berlin : Springer Science & Business Media.
- [23] D. A. Bader, H. Meyerhenke, P. Sanders & D. Wagner. (2011). *10th DIMACS Implementation Challenge-Graph Partitioning and Graph Clustering*. Georgia : Georgia Institute of Technology

김 소 정(SoJeong Kim)

[학사학위]



- 2020년 2월 : 공주대학교 응용수학과 졸업 (이학사)
- 2020년 3월 ~ 현재 : 공주대학교 수학과 (석사)
- 관심분야 : 인공지능, 최적화 알고리즘
- E-Mail : 20200061@smail.kongju.ac.kr

김 찬 수(Chansoo Kim)

[정학위]



- 1991년 : 부산대학교 전산통계학과 졸업
- 1997년 : 부산대학교 통계학과 졸업 (이학박사)
- 2002년 ~ 현재 : 공주대학교 응용수학과 교수
- 2021년 ~ 현재 : 공주대학교 자연과학대학 학장
- 관심분야 : 베이지안 추론, 기계학습, 딥러닝, 빅데이터 분석
- E-Mail : chanskim@kongju.ac.kr

한 근 희(Keunhee Han)

[정학위]



- 1986년 : 건국대학교 물리학과 졸업 (학사)
- 1992년 : Univ. of Central Oklahoma 응용수학과 졸업 (이학석사)
- 1996년 : Univ. of Central Oklahoma 컴퓨터학과 졸업 (이학박사)
- 1996년 ~ 2000년 : 한국전자통신연구원
- 1999년 ~ 2000년 : 미국 NIST 객원연구원
- 2000년 ~ 현재 : 공주대학교 응용수학과 교수
- 관심분야 : 그래프 알고리즘, Genetic Algorithm, Genetic Programming
- E-Mail : kehan@kongju.ac.kr