

# SNA-based Trend Analysis of Naval Ship Maintenance

Jung-Min Yoo\*, Soung-woong Yoon\*, Sang-Hoon Lee\*

## Abstract

Naval ship maintenance generally produces various issues for effective maintenance methods and procedures, because they have been composed by numerous modules and systems, and manual-oriented maintenance needed well-trained technicians who always busy to do many other works. In this paper, we adapt SNA scheme to the service procedure and trends of ROK naval ships' equipments. Various SNA algorithms are deployed which show lots of operating options, and we show analysis results that have enough potential improvement points for the maintainers.

▶ Keyword: SNA, AHP, Centrality, Cohesion, Cosine Similarity, K-Means Clustering

## I. Introduction

해군의 전투력은 함정에 의해 발휘된다. 함정은 많은 종류의 무기체계가 탑재된 복합무기체제로, 해군은 함정이 항상 최상의 전투력을 발휘할 수 있도록 계획적이고 주기적인 정비를 실시해왔다. 그와 동시에 해군은 반복되는 정비를 수행함에 있어, 효과적인 정비방안은 무엇인지 지속적으로 고민해왔으며, 이와 관련된 연구 활동들도 꾸준히 실시되어 왔다.

기존의 연구들은 전체적인 정비 시스템에 대해 초점을 두고 있었다. 함정 정비를 수행하는 정비조직인 정비창에 대해 D/PH/1 모형을 이용하여 최적 정비예약 간격을 결정하는 모형을 제시하는가 하면, 정비 시설인 건선거(Dry Dock)에 대한 모형화 및 시뮬레이션을 통해 효율적인 건선거 규모를 제시하기도 하였고, 시뮬레이션 및 휴리스틱 기법을 이용하여 장비의 수리시간을 최소화 할 수 있는 정비 도구 및 작업 일수를 제시하는 등의 연구가 있었다[1-3].

앞서 언급한 기존의 연구들과 같이 전체적인 시스템에 대해보다 더 효과적인 방안을 연구하고, 적용하는 것도 좋지만, 정비 대상에 대한 이해를 바탕으로 핵심 정비소요를 도출하고, 이 부분에 초점을 두고 정비를 진행하는 것 역시 중요하다.

그러나 기존까지는 특정 정비 대상에 대해 심층적인 특성 파악을 시도하거나, 핵심적인 정비 소요를 도출하려는 노력은 미비하였다. 특히, 해군 함정이라는 특수한 운용환경 속에서 특정

한 장비가 어떠한 정비 패턴을 보이는가에 대한 관련된 연구는 실시된 바가 없다.

본 연구에서는 특정 장비를 샘플로 하여 SNA 기법을 적용을 통해, 해당 장비에 대한 정비 경향을 분석하였고, 이에 대한 이해를 통해 정비자가 집중해야 할 분야를 식별하고자 하였다.

## II. Preliminaries

### 1. Naval maintenance system

해군 함정은 미리 규정된 계획에 따라 통상 연 2회에 걸쳐 일정기간을 정비를 위한 수리기간으로 부여받는다[4]. 함정의 수리기간이 도래하면, 승조원들은 그동안 성능이 저하되었다고 느낀 장비들에 대해 정비부대에 수리 의뢰를 하게 된다.

서론에서 언급하였듯이, 함정은 여러 무기체계가 탑재된 복합무기체계이며, 또한 각각의 무기체계 역시 여러 가지 구성품이 결합된 복합체이다. 따라서 특정 무기체계 또는 장비에 대해 수리 의뢰를 받았을 때, 정비부대 정비자는 장비 전체에 대한 정비를 하기보다는, 기간 및 여건을 고려하여 일부 구성품들에 대해 부분 정비만을 진행하게 된다.

• First Author: Jung-Min Yoo, Corresponding Author: Sang-Hoon Lee

\*Jung-Min Yoo (code\_no1@naver.com), Dept. of Computer Science, Korea National Defense University

\*Soung-Woong Yoon (ysw1209@gmail.com), Dept. of Computer Science, Korea National Defense University

\*Sang-Hoon Lee (hoony@kndu.ac.kr), Professor, Dept. of Computer Science, Korea National Defense University

• Received: 2019. 04. 30, Revised: 2019. 05. 30, Accepted: 2019. 06. 05.

일부 장비의 경우 주기적으로 ‘총 분해 정비’를 수행해야 하는 경우가 있다. 일반적인 경우 총 분해 정비는 함정에 주어지는 수리기간보다 훨씬 많은 시간을 소요로 하며, 이 기간을 단축하기 위해 정비대체장비(M/F)가 확보/운용된다. 즉 총 분해 정비를 수행하는 대신, 총 분해 정비가 수행된 장비로 교체하는 것이다. 이때, 탈거된 장비는 육상에서 정비가 수행되고 또 다른 정비소요를 대체하기 위한 준비 작업을 거친다.

## 2. Related works

### 2.1 SNA(Social Network Analysis)

특정한 관계의 패턴을 가지는 인간이나 집단의 집합을 소셜 네트워크(Social Network)라 한다. SNA는 소셜 네트워크의 구조 또는 형태의 특징을 도출하고, 노드들 간의 관계에 대한 계량적이고 과학적인 해석을 가능하게 하는 분석기법이다[5-6].

SNA는 사회학에서 출발하였으나, 특유의 관계 유형 및 특성 파악의 용이성으로 인해 현재는 경영학, 교육학, 행정학, 컴퓨터공학, 언론정보학 등 다양한 분야로 적용범위가 넓어지고 있다. 또한 고객관계관리(CRM), 모바일 SNS, 전문가 추천 시스템, 정책 및 과학기술 공동연구 네트워크 분석 등 넓은 분야에서 다양한 활용사례를 확인할 수 있다[7].

SNA의 특징은 분석대상이 개체들에게서 측정된 속성 데이터가 아닌, 개체간 파악된 관계데이터라는 점이다[8]. 즉, 링크(Link)에 대한 분석을 통해 노드(Node) 간 관계구조를 파악하는 방법이다.

SNA에서의 데이터 표현은 매트릭스(Matrix)와 그래프(Graph)를 이용한다. SNA에서 노드들 사이의 관계를 표현하는 매트릭스를 인접도 매트릭스(Adjacent Matrix), 또는 소시오 매트릭스(Socio-Matrix)라 하고, 이러한 소시오 매트릭스를 그래프로 나타낸 것을 소시오그램(Sociogram)이라 한다[6]. 이러한 소시오 매트릭스 및 소시오그램의 예는 <그림 1>과 같다.

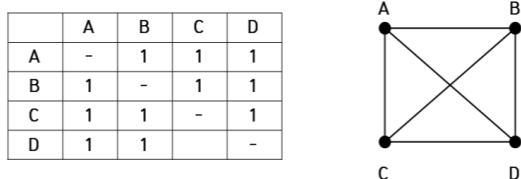


Fig. 1. Representation of socio-matrix and sociogram

### 2.2 AHP(Analytic Hierarchy Process)

AHP는 1971년 Saaty에 의해 제시된, 불확실한 상황이나 다양한 평가기준을 필요로 하는 곳에 쓰이는 의사결정 방법으로, 문제의 분석에 있어 주관적인 판단과 시스템적 접근을 잘 섞어 놓은 문제해결형 의사결정론이다[9].

기본적으로 AHP는 문제에 대해 “분해와 정복”의 원리를 응용한 기법으로 의사결정자로 하여금 의사결정 문제를 일련의 연계된 계층으로 분해하거나 구조화하는 것을 허용한다. AHP는 대안들간 쌍대비교를 수행하고, 모든 대안들을 서열화하고,

의사결정 요소들을 계층으로 구조화함으로써 의사결정 요소들(기준과 대안들)간의 복잡한 관계에 대한 이해의 틀을 제공하는데, 특히 공공정책 분야에 있어 계획과 자원배분, 갈등해소, 예측 등의 분야에 폭 넓게 적용되어 왔다[10].

AHP의 적용은 복잡한 의사결정 과제의 분해, 즉 계층적 구조화에서 시작한다. 계층구조의 맨 위층은 전체적인 목표가 위치하고, 맨 아래층에는 의사결정 대안 또는 선택사항들이 위치한다. 이 두 계층 사이의 중간에는 의사결정의 고려사항이 되는 기준들이 위치한다[9-10].

계층이 구성되면, 전문가들에게 평가항목 간 상대적 중요도의 쌍대 비교를 묻기 위한 설문을 진행한다. 설문을 통해 평가항목들 간 쌍대비교행렬이 얻어지고 나면 고유벡터법을 이용하여 가중치를 도출한다[11]. 또한 쌍대비교행렬에 대해 일관성지수(C.I : Consistency index)를 연산하여 지수가 좋지 않은 경우(일반적으로 0.1 이상) 쌍대비교를 재검토하게 된다[9].

### 2.3 Data Mining on Logistics of National Defense

데이터 마이닝이란 많은 데이터 가운데 숨겨져 있는 유용한 상관관계를 발견하여, 미래에 실행 가능한 정보를 추출하고, 의사결정에 이용하는 과정을 말한다[12].

국방 분야의 데이터 마이닝은 특히 분석을 중심으로 하여 방위산업기술을 분석하거나, 국방 관련 국민 여론을 분석하는 등 다양한 분야에서 활용되고 있다[13]. 특히 국방 군수분야에 있어서는 주로 수리부속의 수요를 예측하는 데 사용되어왔는데, 육군의 경우 인공신경망, 다중선형회귀 등을 활용한 수리부속 수요예측, 공군의 경우 기계학습을 활용한 수요예측, 해군의 경우 랜덤포레스트 모형을 이용한 수요예측 등의 연구가 실시되었다[14-16].

그러나 데이터 마이닝을 국방 정비 분야에 적용한 사례는 없었다. 정비 분야에 있어서도 분야별로 충분히 많은 데이터가 누적되어 있으므로 이에 대한 마이닝을 통해 충분히 의미있는 결과를 이끌어 낼 수 있을 것이다. 본 연구에서는 데이터 마이닝 기법 중 하나인 SNA를 기반으로 그동안 축적된 정비 데이터 가운데 숨겨진 유용한 상관관계를 발견하고자 한다.

## III. The Proposed Scheme

### 1. Study design

본 연구에서는 해군 전투함에 사용되는 추진엔진 중 1개 모델(MTU 20V 956) 23개 개체를 샘플로 하여, 각각의 개체에 대한 정비 실적 데이터 13,682건을 수집하였다. 이렇게 수집된 데이터를 이용하여 총 61개의 하위 구성품을 추출하였으며, 23개 엔진과 61개 하위 구성품을 이용하여 네트워크 데이터를 구성하였고, AHP(Analytic Hierarchy Process) 기법을 사용하여 해군 정비조직의 경험이 반영된 링크 가중치를 산출하여 적용하였다.

구성된 네트워크 데이터를 이용하여 중심성 분석, 유사도 분석, 응집구조 분석, 군집분석 등을 실시하며, 분석 대상 모델의 획득년도와 대상 기간 중의 운전시간(Running Time)과의 상관관계의 유무를 확인한다.

2. Network configuration

2.1 Node data

네트워크를 구성하는 노드들은 ‘엔진’과 같이 같은 유형의 개체들끼리 링크되는 경우가 있고, ‘엔진’과 ‘구성품’처럼 유형이 다른 개체들이 링크되는 경우도 있다. 전자처럼 같은 유형의 노드가 연결된 네트워크를 1-mode network라 하고, 후자를 2-mode network라 한다[6].

본 연구에서는 엔진의 정비경향에 대해 알아보고자 엔진을 하나의 노드로 설정하고, 2.2.1절에서 언급한 것처럼 실질적인 정비가 수행되는 단위인, 엔진의 구성품들을 다른 노드로 설정하여 2-mode network를 구성한다.

보안상의 이유로 불가피하게 노드 표기에 대해 치환을 수행하였는데, 엔진의 경우 해당 모델 전체 개체에 대해 일련번호를 기준으로 문자 ‘G’를, 구성품의 경우 문자 ‘N’을 이용하여 치환 표기한다[17].

2.2 Weight and link data

개체 간 관계, 즉 링크(Link)에 대한 분석을 하는 SNA에서 링크 및 가중치 설정은 큰 의미를 가진다. 본 연구에서는 엔진의 정비 경향을 보다 유의미하게 파악하기 위해 가중치를 적용한다.

정비자들은 엔진의 각 구성품들에 대해 많은 종류의 정비를 수행한다. 예를 들면, 일정시간마다 현재의 성능은 괜찮은지 등을 확인하는 검사, 성능이 저하된 부속에 대한 교체 등이 있을 수 있다. 특정 엔진의 A구성품과 B구성품에 대한 정비횟수가 모두 1회로 동일하더라도, 정비의 종류가 앞서 언급한 대로 검사, 교체 등과 같이 다르다면 그 엔진에 대해 각 구성품이 가지는 중요도의 비중은 같을 수 없다.

본 연구에서의 가중치 적용은 위와 같은 문제를 해결하기 위해 사용되었다. 가중치를 적용하기 위해서는 정비의 종류에 대해 먼저 정의되어야 한다. 그러나 현재까지 해군 내 정비의 종류에 관한 연구 실적은 없으며, 따라서 이번 연구에서는 해군 정비자들의 자문을 얻어 <표 1>과 같이 가중치를 적용하기 위한 정비의 종류를 구분하였다.

Table 1. Types of maintenance

Types	Definition
Inspection	Checking the status of equipment and component, Commissioning for defect identification, etc.
Restoration	Reinstallation of circuit systems, maintenance of software related systems, replacement of consumable items, etc.
Replacement	Replacement of durable items such as new or used parts of components or repair parts

“어떠한 정비가 얼마만큼 더 중요한가?”의 문제는 사실상 객관적인 자료로서는 제공이 불가능하다. 그러나 조직의 경험적인 의사를 반영할 수 있는 AHP 기법을 사용하면 해당 조직이 생각하는 주관적인 중요도, 즉 가중치를 도출해낼 수 있다. 해군의 가장 전문적인 정비조직은 해군 정비창이다. 분야별로 하위 조직이 나뉘어 있으며, 그 중 분석 대상 모델을 정비하는 부서를 대상으로 AHP를 진행하였다.

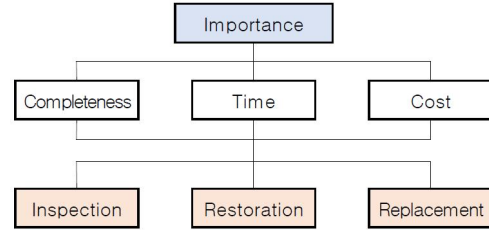


Fig. 2. Structuring the hierarchy

<그림 2>는 본 연구에 있어서 AHP 계층의 구조를 구성한 것이다. 최상층에 궁극적으로 구하고자 하는 가치, 즉 정비중요도가 위치하였고 최하층에는 비교 대상인 3가지 정비 종류가 위치하였다.

최상층과 최하층 사이에는 가중치 산정에 대한 고려요소가 위치하였다. 즉 각 정비 종류에 대한 정비중요도를 평가하기 위해 정비완성도, 소요시간, 정비비용 등 3가지 요소에 대해 비교한다는 의미이며, 각 비교요소의 의미는 <표 2>와 같다.

Table 2. Comparison element

element	Definition
Completeness	Equipment status after maintenance, overall condition such as durability and performance
Time	Comprehensive time require to maintenance, total removal, assembling, commissioning, etc.
Cost	Comprehensive cost to perform maintenance, labor costs, materials costs, etc.

구성된 계층 구조를 바탕으로 정비조직에 대한 조사를 통해 <그림 3>과 같이 쌍대비교행렬을 구성하였다. 행렬의 의미는 C1:정비완성도를 예로 들어, A1:검사는 A2:재생보다 1/5배 중요하다라는 의미이며, A:평가기준의 행렬은 최종목표인 중요도에 대해 각 비교요소들을 비교한 결과이다.

A1 : Inspection, A2 : Restoration, A3 : Replacement

C1 : Completeness				C2 : Time			
	A1	A2	A3		A1	A2	A3
A1	1	1/5	1/7	A1	1	7	4
A2	5	1	1/4	A2	1/7	1	1/6
A3	7	4	1	A3	1/4	6	1

C3 : Cost				A : Comparison Element			
	A1	A2	A3		Completeness	Time	Cost
A1	1	3	7	Completeness	1	6	6
A2	1/3	1	6	Time	1/6	1	1
A3	1/7	1/6	1	Cost	1/6	1	1

Fig. 3. Pair comparison matrix

위와 같이 구성된 쌍대비교행렬에 대해 역승법에 의한 고유치 및 고유벡터 연산을 수행하면 <그림 4>과 같은 결과를 유도할 수 있으며, 2.2.2절에서 언급한 바와 같이 일관성 지수가 모두 0.1 이하이므로 신뢰할 수 있는 조사로 판단하였다.

C1 : Completeness = (0.0692, 0.2437, 0.6871),  $\lambda_{max}$  = 3.1237, C.I. = 0.0619  
 C2 : Time = (0.6795, 0.0644, 0.2561),  $\lambda_{max}$  = 3.1711, C.I. = 0.0855  
 C3 : Cost = (0.6406, 0.2926, 0.0668),  $\lambda_{max}$  = 3.0999, C.I. = 0.0500  
 Comparison element = (0.7500, 0.1250, 0.1250),  $\lambda_{max}$  = 3.0000, C.I. = 0.0000

Fig. 4. Eigenvalue and eigenvector calculation results

<그림 4>에 대해 최종적으로 정비 종류들의 가중치를 도출하면 <그림 5>와 같은 결과를 얻을 수 있는데,

A1. Inspection  
 0.7500 X 0.0692 + 0.1250 X 0.6795 + 0.1250 X 0.6406 = 0.2169  
 A2. Restoration  
 0.7500 X 0.2437 + 0.1250 X 0.0644 + 0.1250 X 0.2926 = 0.2274  
 A3. Replacement  
 0.7500 X 0.6871 + 0.1250 X 0.2561 + 0.1250 X 0.0668 = 0.5557

Fig. 5. Final importance of comparative elements

값이 가장 작은 검사정비 0.2169가 1의 값을 갖도록 조정하면 각각 “검사 : 1, 재생 : 1.0438, 교체 : 2.5618”이라는 최종 가중치를 산출할 수 있다.

해군의 정비 시스템 상 동일한 단위기간 수행되는 각 장비들의 정비횟수는 동일하거나 비슷하다. 이러한 점을 고려하여 각 정비 종류별 가중치에 정비횟수를 곱한 값을 링크데이터로 적용하여 최종 네트워크를 구성하였다.

검사, 재생, 교체정비의 가중치를 각각  $w_1, w_2, w_3$ 라 하고 특정 구성품 N에 대한 종류별 정비 횟수를 각각  $n$ 회라 할 때, 링크 가중치는 아래와 같이 표현할 수 있다.

$$\sum_{i=1}^3 \sum_{j=1}^n w_i N_j$$

이렇게 구성된 네트워크는 “엔진-구성품” 간 정비 중요도에 대한 관계를 해당 정비실적과 정비부서의 주관적인 가중치를 이용하여 표현한 것으로, 해군 정비부서가 엔진에 대해 여러 구성품을 상대적으로 얼마나 중요하게 여기는지에 대한 수치를 나타낸다.

### 3. Core components extraction

노드 중심성은 네트워크의 어느 노드가 중요한가를 나타내는, SNA에 있어 널리 쓰이는 개념 중 하나이다[18]. 본 연구에서는 엔진의 여러 구성품들 중 핵심 구성품을 추출하기 위해 Bonacich(1972)에 의해 정의된 ‘Eigenvector Centrality’를 사용한다[19].

Eigenvector Centrality의 기본 개념은, 가장 중심이 되는 노드는 인접 노드와의 연결 관계가 많은 노드가 아니라 중심성

이 높은 노드와 연결 관계가 많은 노드라는 것이다[18].

Bonacich는 아래와 같이 노드  $v_i$ 의 중심성  $c(v_i)$ 를 인접한 노드들의 중심성간 합의 양의 배수로 정의하였는데,

$$\lambda c(v_i) = \sum_{j=1}^n a_{ij} c(v_j) \quad \forall i$$

이에 대해  $c$ 를 이용한 행렬 표기법을 사용하면,  $\lambda c = Ac$ 와 같이 표기할 수 있고, 이 방정식은 행렬  $A$ 에 대한 고유값 연산과 고유벡터 연산으로 풀이할 수 있다[19].

본 연구에 있어서 위와 같은 Eigenvector Centrality를 이용하여 구해지는 핵심 구성품들은 “중심성이 높은 엔진들과 높은 연결 관계를 가지는 구성품”의 의미를 가진다.

### 4. Ranking of engine by maintenance score

엔진의 정비 순위 평가는 최초 구성된 네트워크의 엔진 노드들에 대한 노드 연결정도(degree) 비교, 3.3절에서 언급한 핵심 구성품들에 대한 네트워크를 별도 추출한 후 이에 대한 degree 비교, 마지막으로 핵심 구성품 추출 시와 동일하게 Eigenvector centrality를 이용한 엔진들의 중심성 순위 비교 순서로 진행한다.

$d(n_i)$ 로 표현하는 노드의 연결정도, degree는 그 노드와 연결된 링크 수를 의미한다. 동일하게, 그 노드와 연결된 노드의 수를 의미하기도 하며 가중치가 있는 네트워크의 경우 해당 노드가 가진 모든 링크의 가중치를 더한 값이 되기도 한다[6].

이러한 비교/평가를 통해, 먼저 모든 정비가 누적된 순위를 확인을 통해 엔진들의 도입연도 또는 운전시간과 같은 속성들이 순위에 미치는 영향들을 파악해 보고, 이후 핵심 구성품들의 정비순위 분석을 통해 가장 잦은 정비가 가장 많이 일어나는 엔진 등을 찾아낼 수 있다. 또한 마지막으로 전체 구성품들에 대해 전체적인 정비 중심성이 높은 엔진을 찾아낼 수 있다.

### 5. Similarity network creation analysis

#### 5.1 Creation of similarity network

2-mode network 데이터의 속성을 이용하여, 유사도 또는 거리를 계산하고 이를 1-mode network 데이터로 변환하여 분석하는 방법은 SNA의 대표적인 분석방법 중 하나이다[20].

유사도 계산에는 여러 가지 방법들이 제시되어 있는데, 본 연구에서는 대표적인 변환 방법인 코사인 유사도를 이용하기로 한다[21]. 코사인 유사도는 아래와 같이 표현할 수 있다[22].

$$\sigma_{\cos} = \frac{|E_i \cap E_j|}{\sqrt{|E_i| |E_j|}}$$

즉, 코사인 유사도는 두 벡터의 내적을 해당 유향성분으로 나눈 값으로 풀이될 수 있다. 이를 본 연구에 적용하자면, 먼저 엔진-구성품 간 정비자료 네트워크를 이용하여 각 구성품들에 대한 엔진의 벡터를 추출한다.

$$G_n = (N_1, N_2, N_3, \dots, N_n)$$

본 연구에서 다루는 구성품의 종류는 61종이므로 61차원의 엔진 벡터가 생성되고, 생성된 엔진 벡터들의 코사인 유사도, 즉  $\cos\theta$  값을 아래와 같이 socio-matrix로 표현할 수 있다.

/	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	...
G <sub>1</sub>	-	$\sigma_{G1,G2}$	$\sigma_{G1,G3}$	...
G <sub>2</sub>	$\sigma_{G2,G1}$	-	$\sigma_{G2,G3}$	...
G <sub>3</sub>	$\sigma_{G3,G1}$	$\sigma_{G3,G2}$	-	...
...	...	...	...	-

Fig. 6. Cosine similarity socio-Matrix

위의 절차와 같이 생성된 네트워크는 각 구성품들에 대한 정비경향의 유사성에 기반한 유사도를 링크로 하는, 엔진 노드끼리의 1-mode network가 된다.

## 5.2 Analysis of cohesion structure

엔진 노드끼리의 유사도 네트워크는 각 노드 벡터에 대해 다른 모든 노드 벡터를 연산하므로, 결과적으로 모든 노드가 다른 모든 노드에 연결된 형태를 갖는다. 즉 유사도가 아주 낮은 Node 사이에도 Link가 생성되게 된다.

본 연구에서 확인하고자 하는 바는 어떠한 엔진 노드끼리 더 유사한가에 대한 내용이므로 위에서 언급한 유사도가 낮은 Link는 고려할 필요가 없다.

코사인 유사도는 최소 0, 최대 1의 값을 갖는다. 본 연구에서는 90% 이상 유사한 Link만을 남기고 이 외 링크는 삭제되도록 Link Threshold를 조절하였으며, 이를 통해 가장 핵심적인 특징을 나타내는 네트워크를 추출할 수 있다.

네트워크에서의 응집구조를 확인하고 찾아내는 방법은 그동안 많이 연구되어온 분야이며, 그 방법도 다양하다[6]. 이번 연구에서는 ‘커뮤니티’라는 응집구조를 사용한다. 커뮤니티란 그룹 내 링크가 그룹 간 링크보다 많도록 분류된 노드들의 집합이며, 그룹 내 노드끼리는 강하게 연결되어 있고 그룹 외 노드와는 느슨하게 연결되어 있는 구조이다[23]. 또한 하나의 노드는 반드시 하나의 그룹에만 속하게 된다.

물론 이러한 커뮤니티 구조를 발견하기 위한 방법들도 다양하게 제시되어 있다[23]. 이번 연구에서는 Newman이 제안한, Modularity를 극대화하는 알고리즘을 사용한다[24].

Modularity란 Clustering이 얼마나 잘 되었는지를 측정하는 지표로서 최대값 1에 가까울수록 좋은 Cluster라는 의미이며, 최소 0.3이상일 때 유의미한 Cluster라 할 수 있다[24]. 이는 아래와 같이 “현재 그룹 내 링크수” - “랜덤 네트워크에서 기대되는 그룹 내 링크 수”의 의미를 갖는 수식으로 나타낼 수 있는데, 최초 모든 링크가 분리된 노드에서 시작하여 링크를 하나씩 추가하면서 Modularity가 가장 커지는 구조를 찾는 것이 알고리즘의 핵심이다[24].

$$Q = \frac{1}{2m} \sum_{vw} [A_{vw} - \frac{k_v k_w}{2m}] \delta(c_v, c_w)$$

$m$  : number of links  
 $k_v$  : degree of node  $v$   
 $\delta(i, j) = 1, \text{ if } i = j,$   
 $0 \text{ otherwise}$   
 $A_{vw} : 1, \text{ if } v \text{ and } w \text{ are connected}$   
 $0, \text{ otherwise}$

## 5.3 Ego-Network extraction of engine

2.1절에서 언급한 바와 같이 해군은 정비시간을 단축하기 위해 정비대체장비(M/F)라는 여분의 장비를 운용중이다. 이러한 경우, 향후 정비 소요를 감소시킬 목적으로 유사도 네트워크의 Ego-Network를 이용할 수 있다.

Ego-Network란 해당 노드 및 그 노드와 연결된 노드끼리만 구성된 네트워크를 의미한다. 즉 노드 1개에 대한 Adjacent Matrix로 표현될 수 있으며, 앞서 생성된 유사도 네트워크에서 각 엔진별 유사도의 Ego-Network를 추출할 수 있다[6].

본 연구에 사용된 샘플(엔진)은 총 2대가 함정에 탑재되는 타입으로, 좌/우현 엔진이 구분되어 있다. 긴급/고장 복구 등의 사유로 인해 1대의 엔진만을 교체해야하는 상황이라 가정하자. 그렇다면 이러한 정비경향의 유사도를 이용한 Ego-Network에서 링크의 순위가 가장 높은 이웃 노드를 교체할 엔진으로 선택하는 것이 향후 유사 정비를 통합 실시할 수 있다는 점에서 유익할 수 있다. 동일하지는 않겠지만 다른 엔진 Pair에 비해 향후 정비경향이 더 비슷할 것으로 기대되기 때문이다.

## 6. Engine clustering and profiling

앞서 3.5.2절에서는 엔진 정비경향의 유사도를 이용하여 응집구조를 확인하는 방법을 제시하였다. 이와 조금 다르게 K-means clustering을 이용하면 구성품의 계통(System)별로 엔진을 군집분석할 수 있다.

이번 연구에서 사용하는 구성품 총 61종은 나뉘는 계통별로 분류할 수 있다. 예를 들면 청수펌프는 청수계통, 윤활유냉각기는 윤활계통 등으로 분류할 수 있으며 이렇게 분류한 결과 총 10개 계통을 얻을 수 있었다.

분류된 계통을 이용하여 “엔진-계통” 네트워크를 구성하면 각 엔진별 계통에 대한 정비경향을 링크로 얻을 수 있는데, 이 링크 값을 해당 엔진의 모든 링크의 합으로 나누어 주면 해당 엔진에서 해당 계통이 차지하는 정비비율을 얻을 수 있다.

K-means 알고리즘은 개체의 속성을 이용하여 군집화하는 대표적인 알고리즘이다. 최초 사용자가 Cluster의 수를 설정하면 초기 k값이 정해지고 그룹 평균 k와 노드간 거리의 분산을 최소화하는 방식으로 동작한다[25].

엔진에서 해당 계통이 차지하는 정비비율을 속성으로 하고, 이를 K-Means Clustering하면 3.5.2절에서 설계한 응집구조 분석과는 다른 결과를 얻을 수 있고, 이러한 비율을 통해 엔진에 대한 프로파일링이 가능하다.

## IV. Analysis Results

### 1. Result of core components extraction

엔진-구성품 네트워크를 이용하여 Eigenvector Centrality 0.1 이상의 상위 20개 구성품을 추출한 결과는 <표 3>과 같다.

Table 3. Core components

Rank	Comp'	Centrality	Rank	Comp'	Centrality
1	N51	0.3152	11	N34	0.1963
2	N12	0.3119	12	N11	0.1726
3	N2	0.3021	13	N24	0.1662
4	N7	0.2998	14	N16	0.1612
5	N13	0.2825	15	N33	0.1455
6	N1	0.2605	16	N29	0.1418
7	N9	0.2320	17	N27	0.1280
8	N15	0.2311	18	N5	0.1219
9	N6	0.2213	19	N3	0.1197
10	N52	0.1987	20	N31	0.1107

중심성 수치는 N<sub>51</sub>이 0.3512로 가장 높게 나타났고, N<sub>55</sub>가 0.0027로 가장 낮게 나타났다. 평균 중심성 수치는 0.0864였으며, 전체 중심성의 합계는 5.2746, 이중 <표 3>의 상위 20개 구성품의 중심성 합계는 4.1190으로 전체의 78%를 차지하는 높은 비율을 보였다.

즉, <표 3>에 제시된 핵심구성품에 의해 전체적인 엔진의 정비점수가 결정된다고 분석할 수 있으며, 전체 엔진들 중 가장 정비가 빈번하게 수행되는 구성품들이라 할 수 있다.

### 2. Comparison of engines by maintenance score

<표 4>는 3.4절에서 설계한 바와 같이 엔진 노드들에 대해 정비순위를 비교한 결과이다. 제일 왼쪽부터 전체 구성품에 대한 연결정도의 합(Sum of degree), 4.1절에서 추출한 핵심 구성품들에 대한 연결정도의 합, 마지막으로 엔진 노드들에 대한 Eigenvector Centrality를 순위대로 나열하였다.

Table 4. Maintenance ranking of engines

Rank	Basic		By core components				Eigenvector	
	Node	Score	Node	Obtain	R.time	Score	Node	Score
1	G10	294.8	G10	2000	9,508	208.5	G10	0.2974
2	G7	257.9	G7	1998	11,603	197.6	G15	0.2722
3	G15	242	G15	2003	13,774	193.7	G7	0.2683
4	G12	236.3	G12	2000	8,721	178.5	G3	0.2542
5	G3	230	G3	2004	9,314	177.9	G12	0.2528
6	G14	218.8	G17	2004	10,662	172.1	G17	0.2441
7	G5	213.1	G5	1998	9,090	169.8	G14	0.2387
8	G17	211.4	G21	2006	10,239	166.5	G5	0.2354
9	G18	210.1	G14	2003	10,813	166.1	G18	0.2343
10	G23	203.4	G18	2004	12,032	162.8	G21	0.2233
11	G19	202.1	G4	1999	6,115	161.8	G9	0.2230
12	G21	201.7	G23	2008	11,222	159.0	G23	0.2182
13	G4	200	G9	2007	6,913	158.3	G19	0.2159
14	G9	192.9	G13	2007	3,874	153.4	G4	0.2138
15	G8	188.5	G8	2005	10,932	152.2	G13	0.2087
16	G13	187.6	G19	2008	12,570	150.5	G8	0.2034
17	G6	171.8	G6	1999	10,847	139.6	G6	0.1916
18	G22	152.5	G22	2006	9,117	125.5	G22	0.1713
19	G2	134.5	G2	2005	5,586	98.4	G2	0.1402

핵심 구성품들에 의한 정비점수 순위에는 해당 엔진의 획득년도(Obatin), 분석 기간 중 운전시간(Running Time)을 같이 표기하여 상관관계를 확인하고자 하였다.

<표 4>에서도 전체 구성품에 의한 정비점수의 순위와 핵심 구성품들에 의한 정비점수 순위는 전 절에서와 비슷한 경향을 보이고 있으며 그 차이가 거의 없는 수준이다. 즉, 추출된 핵심 구성품들에 의해 정비순위 역시 결정된다고 분석할 수 있다.

단순하게 획득 후 운용연수가 길수록 노후된 장비이며, 유지보수에 더 많은 노력이 필요하다고 생각할 수 있으나, <표 4>가 보여주는 결과는 다르다. G<sub>6</sub>와 같이 오래 전(1999년) 획득한 장비의 정비점수가 총 19개 엔진 중 17위로 낮은가 하면, G<sub>10</sub>은 G<sub>6</sub>보다 1년 늦은 2000년 획득되었으나 정비점수는 1위로 가장 높았다. 전체적으로 1위에서부터 19위까지 획득년도는 고르게 분포되어 있으며, 따라서 획득년도가 정비순위 결정에 영향을 끼쳤다고 판단할 수 없다.

운전시간의 경우도 동일하다. 단순히 더 많이 운전을 한 장비의 정비점수가 높을 수 있다 생각되나 각각 14, 15위를 차지하는 G<sub>13</sub>과 G<sub>8</sub>에서의 케이스 등과 같이 운전시간이 정비순위에 영향을 미쳤다고 판단할 근거는 없다. 따라서 운전시간이 정비순위에 영향을 미치지 않는다는 점이 확인되었다.

### 3. Analysis result of engine cohesion structure

코사인 유사도를 관계로 하여 엔진끼리의 유사도 네트워크를 생성한 뒤 3.5.2절에서 설계한 것처럼 응집구조 분석을 진행하였다. 앞서 언급한 것처럼 모든 노드가 서로에게 모두 연결된 형태의 유사도 네트워크가 생성되므로 그 특징을 나타낼 수 있도록 Link Threshold를 0.9 이상으로 조정하여 <그림 7>과 같은 네트워크를 얻었다.

생성된 유사도 네트워크의 Link값은 코사인 유사도로서, 최소 0과 최대 1의 값을 가지므로 0.9 이상을 추출한 결과는 유사도 90% 이상인 링크만을 남긴 결과와 동일하다. 또한 정비패턴이 90% 이상 동일한 엔진끼리의 연결만 남겨두게 된다.



Fig. 7. Similarity network

<그림 7>의 네트워크에서 고립(Isolated) 노드인 G<sub>7</sub>, G<sub>2</sub>, G<sub>8</sub>, G<sub>22</sub>를 제외하고 나머지 연결된 노드들을 이용하여 응집구조 분석을 진행한 결과는 <그림 8>과 같다.

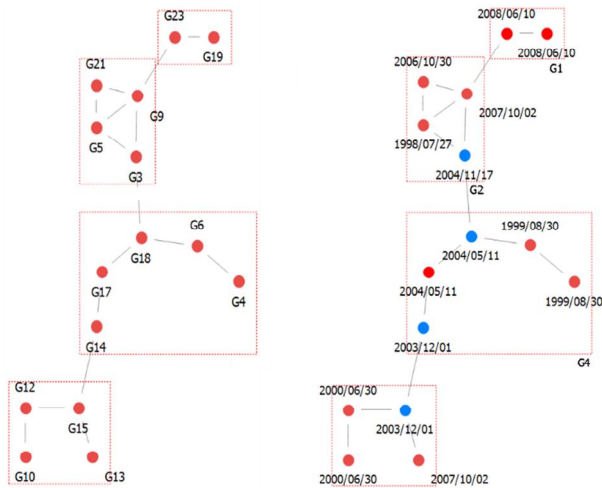


Fig. 8. Cohesion analysis

Member가 각각 2 ~ 5개인 4개 그룹으로 응집되어 있음을 확인하였으며, 이때 Modularity는 0.518로써 유의미한 Cluster라 판단할 수 있다[24].

4.2절과 같이 운전시간, 획득년도 등 두 가지 요인을 적용하였을 때, 엔진의 획득년도가 <그림 8>과 같이 유의미한 패턴을 보였다. 그룹 내 엔진들은 도입시기가 비슷하거나 같은 경향을 보이며 엔진 간 링크 역시 도입시기가 비슷한 엔진들끼리 연결되는 경향을 보인다. 또한 파란색으로 표기한 노드들(G<sub>3</sub>, G<sub>18</sub>, G<sub>14</sub>, G<sub>15</sub>)처럼 그룹 외로 연결되는 링크 또한 도입시기가 비슷하거나 같은 엔진들끼리의 링크였다.

해군의 정비자들은 함정의 두 대 엔진 중 하나의 엔진에 정비요소가 있을 때, 가급적 나머지 엔진에도 예방정비 차원의 동일 정비를 실시하는 경향이 있다. 즉, 동일 함정에 탑재된 엔진끼리의 정비경향이 시간이 갈수록 비슷해질 수 있는 것이다.

그러나 서로 같은 그룹으로 분류된 G<sub>10</sub>과 G<sub>12</sub>, 그리고 다른 그룹으로 분류된 G<sub>21</sub>의 예를 자세히 살펴보면 흥미로운 결과를 확인할 수 있다. G<sub>10</sub>과 G<sub>12</sub>가 동일한 함정에 탑재된 기간은 4년, G<sub>10</sub>과 G<sub>21</sub>이 동일한 함정에 탑재된 기간은 5년이였다. 그러나 링크는 G<sub>10</sub>과 G<sub>21</sub>사이가 아니라 G<sub>10</sub>과 G<sub>12</sub>사이에 있었으며, 이는 동일함정 탑재기간 보다 획득시기가 정비경향 유사성에 더 영향이 있다고 분석할 수 있는 근거가 된다.

즉, 동일하나 비슷한 시기에 획득된 엔진들은 비슷한 정비경향을 갖는다는 것이 확인되었으며, 이를 통해 다른 함정에 탑재되어 있더라도 획득시기가 비슷한 엔진이라면 해당 엔진의 정비 및 결함을 참고해야 한다고 분석할 수 있다.

4. Result of ego-network extraction

생성된 유사도 네트워크를 이용하여 Ego-Network를 추출한 결과는 <표 5>와 같다.

3.5.3절에서 언급한 바와 같이 <표 5>와 같은 Ego-Network 도표는 M/F pair를 찾기 위해 사용할 수 있다. 예를 들면 이미 G<sub>2</sub>가 탑재되어 있는 함정에서, 나머지 한 대 엔진에 대한 M/F 교체를 계획하고 있다면 <표 5>가 보여주는 G<sub>4</sub>, G<sub>18</sub> 등과 같은 순위대로 유사성에 기반하여 선택할 수 있다.

Table 5. Ego-network of engine node

Rank	G2	G3	G4	G5	G6	G7	G8	G9	...
1	G4	G5	G6	G21	G4	G8	G7	G5	...
2	G18	G9	G18	G9	G18	G15	G6	G3	...
3	G5	G18	G2	G3	G14	G10		G21	...
4	G14	G22	G17	G23	G17	G5		G23	...
5	G9	G21	G21	G10	G7	G13		G13	...
6	G10	G17	G14	G7	G15	G12		G22	...
7	G19	G6	G12	G13	G3	G21		G15	...
8	G15	G23	G3	G22	G8	G6		G18	...
9		G12	G23	G12	G22	G3		G19	...
10		G7	G15	G15	G10	G4		G17	...
11		G15	G7	G19	G5			G12	...
12		G4		G18	G9			G14	...
13		G13		G2	G12			G6	...
14				G6				G2	...
15				G17					...

이러한 선택은 향후 발생할 정비의 유사성을 최대한 높이는 방향과 같아지므로, 두 대 엔진의 동시정비 가능성을 최대한 높임으로서 비교적 높은 정비효과를 기대할 수 있다.

5. Result of engine clustering

3.6절에서 언급한 바와 같이 엔진들을 구성품의 계통별 정비 실적에 따라 군집분류를 위해 먼저 계통분류를 처리한 내용은 <표 6>과 같다.

Table 6. Component system

Hydraulic	Inhalation	Power	Fresh water	Air	Lubrication	Sea water	Fuel	Circuit	Exhaust
N1	N2	N4	N5	N7	N8	N10	N12	N13	N16
N3	N23	N22	N6	N14	N9	N11	N29	N34	N24
		N27	N15	N19	N18	N20	N31	N45	
		N28	N17	N21	N32	N25	N37	N56	
		N30	N26	N33	N44	N35	N39	N57	
		N38	N41	N36	N50	N51	N43	N58	
		N40	N42	N53		N52	N46	N59	
		N47	N48	N60			N54		
		N55	N49						
		N61							

이렇게 분류된 계통을 이용하여 엔진-계통 간 네트워크를 구성하였다. 이때의 링크 데이터의 계통별 분포 비율은 <그림 9>과 같이 나타났다. 청수와 공기계통의 비율이 15% 이상으로 가장 높았고 가장 낮은 비율들은 약 6~7%를 차지하고 있다.

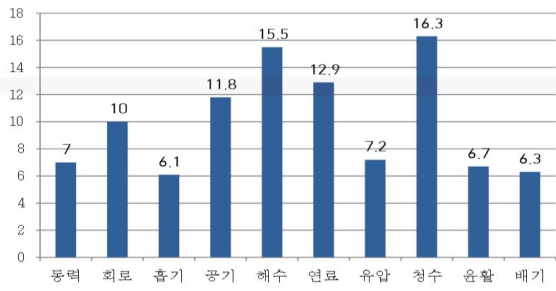


Fig. 9. Percentage of systems

위와 같은 전체 비율과 비슷하게, 엔진 노드 각각에 대한 계통의 비율도 네트워크 링크를 통해 표현될 수 있다. 링크를 통해 표현한 엔진별/계통별 비율은 <그림 10>과 같다.

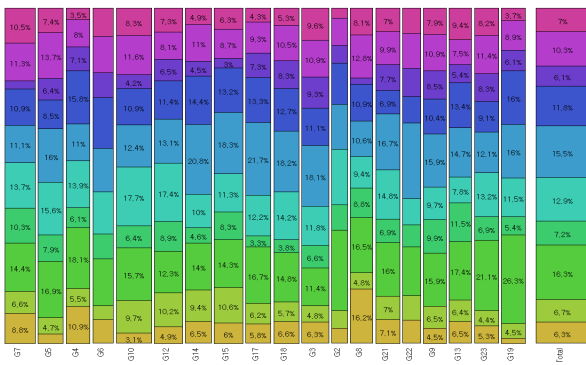


Fig. 10. System ratio of engines

<그림 8>의 링크의 비율은 각 엔진 노드들의 속성으로 사용할 수 있다. 즉, 각 엔진들에 대해 비교/군집화를 처리할 수 있는 10개 속성이 마련된 것이다.

K-Means Clustering은 분류하고자 하는 Cluster의 개수를 최초 설정해 주어야 한다[25]. 이번 연구에서는 각 Cluster의 특징이 가장 뚜렷이 구분되는 4개의 Cluster 개수를 사용하였다. K-Means Clustering을 실시한 결과는 <그림 11>과 같다.

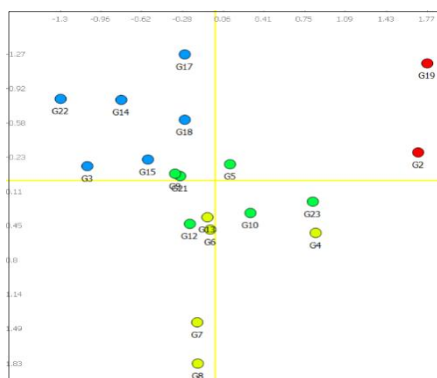


Fig. 11. K-Means clustering result

<표 7>은 Clustering 결과를 상세히 나타낸 것이다. 이를 통해 상대적으로 각 그룹별 비중이 높은 계통과 낮은 계통을 찾

을 수 있는데, 예를 들면 1번 그룹은 타 그룹에 비해 청수, 공기 계통의 비중이 높으며 배기, 윤활, 회로, 동력 계통의 비중이 낮게 나타나고 있다.

Table 7. Clustering result

Cluster	Exhaust	Lubrication	Fresh water	Hydraulic	Fuel
	Sea water	Air	Inhalation	Circuit	Power
1	0.03	0.049	0.251	0.065	0.14
	0.147	0.141	0.057	0.086	0.034
2	0.099	0.055	0.161	0.088	0.115
	0.121	0.132	0.049	0.101	0.08
3	0.046	0.071	0.163	0.078	0.147
	0.144	0.095	0.069	0.109	0.077
4	0.064	0.073	0.143	0.055	0.112
	0.199	0.121	0.068	0.102	0.063

위 결과를 보다 이해하기 쉽도록 <표 8>과 같이 나타낼 수 있다. 이렇게 나타낸 표는 엔진별 취약 계통은 무엇인지와 이를 이용해, 보다 집중 정비가 필요한 부분은 어디인지를 나타내다 고 할 수 있다.

Table 8. Details of clusters

No.	Engine	Color (Fig.9)	High percentage	Low percentage
1	G2, 19	Red	Fresh water, Air	Exhaust, Lubrication, Circuit, Power
2	G4, 6, 7, 8, 13	Yellow	Exhaust, Hydraulic	Sea water, Inhalation
3	G5, 9, 10, 12, G21, 23	Green	Fuel, Circuit	Air
4	G3, 14, 15, G17, 18, 22	Blue	Lubrication, Sea water	Fresh water, Hydraulic, Fuel

## V. Conclusion and Future Works

이번 연구에서는 효과적인 정비를 수행하기 위한 방법을 찾고자 SNA를 이용하여 함정장비에 대한 정비경향을 분석해 보았다. 기존의 장비 시스템 전체에 집중하는 관점에서 벗어나 각 개체장비에 대한 경향 및 패턴 등에 대해 파악하고자 하였다. 분석 결과 다음과 같은 결론을 얻었다.

첫째, 핵심구성품 20종을 식별하였으며, 이 구성품들이 전체 정비경향을 결정할 수 있을 만큼의 비율을 차지하는 것을 확인하였다. 정비자들은 이 핵심구성품들에 대한 정비를 우선하여 보다 효과적인 정비계획을 수립할 수 있을 것이다.

둘째, 엔진의 정비순위는 도입년도와 장비 운전시간과는 무관한 관계를 보였다. 이러한 결과는 장비의 성능적인 측면에 대해서 언급할 수는 없으나, 정비비용 측면에서는 노후도와 큰 상관관계가 없는 것 분석할 수 있다.

셋째, 유사도 네트워크를 이용한 응집구조 분석은 비록 다른 함정에 탑재되어 있던 기간이 길더라도, 동일하거나 비슷한 시기에 획득된 엔진들은 비슷한 정비경향을 갖는다는 결과를 보였다. 또한 이러한 유사도 네트워크는 Ego-Network 추출을



통해 이미 1대의 엔진이 선택된 상황에서 M/F Pair를 찾는 데 유용하게 쓰일 수 있음을 확인하였다.

넷째, K-Means를 통한 군집분석은 각 엔진 사용자와 전문 정비자들에게 개체장비 각각에 대해 어떠한 부분에 보다 집중하여 정비를 수행해야 하는지에 대한 결과를 나타내고 있다.

연구의 샘플로 사용한 대상은 추진엔진이었으나, 이번 연구에서 사용한 기법들은 추진엔진뿐만 아니라 다른 종류의 함정 장비에도 적용 가능할 것이다. 이번 연구에서 분석된 결과를 이용하여 이전보다는 비교적 효과적인 정비가 가능하리라 판단한다. 서두에서 언급하였듯 효과적인 방법 중 하나는 그 대상에 대해 보다 자세히 파악하는 것이기 때문이다.

SNA를 이용한 데이터마이닝은 앞서 기술한 바와 같이 각 개체장비의 특성을 파악하는데 유리하게 적용할 수 있음을 또한 확인하였다. 다품종을 기반으로 한 기존의 수리부속 예측 등 국방분야 데이터마이닝 적용 사례와는 다른 측면으로, 각각 개체장비에 대한 유용한 상관관계 발견은 SNA 기법을 적용하여 가능하였다.

향후에는 동일한 대상을 이용하여 가중치를 전력 공백의 값, 정비시간, 부품 단가 등으로 하는 고장 복구 네트워크를 분석할 예정이다. 이번 연구의 결과와 향후 연구의 결과가 더해진다면 해군 정비분야에 보다 유의미한 결과를 전달할 수 있으리라 판단한다.

## REFERENCES

- [1] Go, Jae-woo, "A Study on the Optimal Appointment Scheduling for the Ship Maintenance with Queueing System with Scheduled Arrivals," Journal of The Korean Operations Research and Management Science Society, Vol. 38, pp. 13-22, 2013.
- [2] Ok, Kyung-chan, "A Study on an Efficient Size Docks for Warship Maintenance using Queueing Problem," Journal of the Korean Institute of Industrial Engineers, Vol. 40, pp. 428-434, 2014.
- [3] Kim, Kyung-rock, "Simulation Analysis to optimize the management of Military Maintenance Facility," Journal of the Korean Academia-Industrial cooperation Society, Vol. 15, pp. 2724-2731, 2014.
- [4] ROK NAVY, "Ship's Maintenance System," pp. 9-10, 2018
- [5] Jung, Chi-young, "A Study on the Measurement Model for NCW's Synergistic Effectiveness using SNA," Korea National Defense University, 2012
- [6] Wasserman, S. and Faust, K., "Social Network Analysis : Methods and Applications," Cambridge University-Press, 1994.
- [7] Kim, Sang-kook, "Social Network Analysis(SNA) and Industrial Engineering," IE-magazine, Vol. 18, No. 1, pp. 24-32. 2011.
- [8] Lee, Soo-Sang, "Network Analysis Methods," Nonhyung, 2012
- [9] kinoshita, "Strategic Decision Making Techniques," Chung-ram, 2012
- [10] kim, Yoon-ju, "A Comparison of Weight Elicitation Techniques: Focusing on AHP, JA, and SW," National Knowledge Information System, Vol. 21, No. 1, pp. 5-33, 2007.
- [11] Lee, Young-ju, "Principles of AHP analysis," research report, EWHA womans university, May 2001.
- [12] Data mining, <https://terms.naver.com/entry.nhn?docId=1179774&cid=40942&categoryId=32841>
- [13] Son, Chang-ho, "Study for Analyzing Defense Industry Technology using Datamining technique - Patent Analysis Approach -," Journal of the Korea Academia-Industrial cooperation Society, Vol. 19, No. 10, pp. 101-107, October, 2018
- [14] Kim, Hyung-tae, "Data Mining based army repair parts demand forecast," Journal of the Korean Data & Information Science Society, Vol. 30, No. 2, pp. 429-444, March, 2019
- [15] Jung, Won-hee, "Machine Learning-based System for Predicting the Demand for Spare parts and Maintenance of Air Force Equipment," Journal of Korean Institute of Information Scientists and Engineers, Vol. 36, No. 8, pp. 21-26, August, 2018.
- [16] Yoon, Hyun-min, "A Study on forecasting spare parts demand using data mining," Korea National Defense University, 2017
- [17] Yoo, Jung-min, et.al, "Deducting Core Parts of ROK Naval Ship's Engine based on SNA," Proceedings of KSCI Conference, Vol. 27, No. 1, pp. 419-422, January, 2019.
- [18] B.Ruhnau, "Eigenvector-centrality - a node-centrality?," Social-Networks, Vol. 22, pp. 357-365, 2000
- [19] Bonacich P. "Factoring and Weighting Approaches to status scores and clique identification," Journal of Mathematical Sociology, Vol. 2, pp. 113-120, 1972.
- [20] Borgatti, S. P., & Everett, M. G. "Network analysis of 2-mode data," Social Networks, Vol. 19, No. 3, pp. 243-269, 1997
- [21] G. Salton, "Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer," Addison-Wesley, Reading, MA, 1989.
- [22] Leicht, E. A., Holme, P., & Newman, M. E. J. "Vertex similarity in networks." Physical Review E, Vol. 73, No. 2, Feb, 2006
- [23] M. Girvan & Newman, "Community structure in social

and biological networks," PNAS, Vol. 99, No. 12, pp. 7821-7826, June, 2002

- [24] Clauset, A. et al, "Finding community structure in very large network," Physical Review, Vol. 70, No. 6, 06111, 2004
- [25] Hartigan, J. A. et al, "Algorithm AS 136: A K-Means Clustering Algorithm," Journal of the Royal Statistical Society, Series C 28, pp. 100-108, 1979. Introduction

## Authors



Jung-Min Yoo received the B.S degree in Computer Science from Korea Naval Academy in 2008 and He is a candidate for M.S. degree in Computer Science from Korea National Defense University. He is interested in Big data, Social

Network Analysis and knowledge discovery.



Soung-Woong Yoon received the B.S. degree in Engineering from Hanyang University in 1992 and M.S. degree in Computer Science and Engineering from Korea National Defense University (KNDU), Korea, in 2004, respectively.

Mr. Yoon completed Ph.D course of the Department of Social Informatics at Kyoto University, Japan, in 2014. He is interested in knowledge discovery, information retrieval, big data and Social Network Analysis.



Sang-Hoon Lee received the B.S. degree in Electronic Engineering from Sung Kyun Kwan University, M.S. degree in Computer Science from Yon Sei University, and Ph.D. degree in Information Engineering from Kyoto

University, Japan, in 1978, 1987 and 1997, respectively. Dr. Lee joined the faculty of the Department of Computer Science at Korea National Defense University, in 2000. He is currently a Professor in the Department of Computer Science, Korea National Defense University. He is interested in DataBase, Big Data, Social Network Analysis, and Information search.