JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 24, No. 7: 914~924, Jul. 2020

온라인 소셜 네트워크 생성 모델

이강원*

On-Line Social Network Generation Model

Kang-Won Lee*

*Professor, Department of Industrial Engineering, Seoul National University of Science and Technology, Seoul, 01811 Korea

요 약

본 연구에서는 소셜 네트워크를 생성 할 수 있는 인공적인 네트워크 발생 모델을 제안 하였다. 본 연구에서 제안한 발생 모델은 온라인 소셜 네트워크의 특징인 Small-World 성질과 Scale-Free 성질을 단순하게 표현하는 것에서 벗어나 모델의 두 파라메터를 적절히 조절함으로서 사용자가 원하는 다양한 위상 특성치 값들을 나타내 줄 수 있도록 하였다. 이를 위해 Preferential Attachment의 세기를 조정 할 수 있도록 파라메터 K와 군집화 계수를 적절하게 조정 할 수 있도록 파라메터 P를 도입하였다. K가 0에서 10 그리고 P가 0.3에서 0.5 사이의 조합이나 K = 0과 P = 0.9를 이용하면 소셜 네트워크의 위상적 성질을 보유하는 인공적인 네트워크를 생성할 수 있다. 이 조합 하에서는 Small-World 성질과 Scale-Free 성질이 잘 나타난다. 노드차수 분포는 Power-Law를 따른다. 또한 군집화 계수 0.130 ~ 0.238, 평균 최단거리 5.641 ~ 5.985로 나타났다. 또한 네트워크의 크기를 노드 5,000개에서 10,000개로 증가시켜도 소셜 네트워크 성질을 그대로 유지하는 것으로 나타났다.

Abstract

In this study we developed artificial network generation model, which can generate on-line social network. The suggested model can represent not only scale-free and small-world properties, but also can produce networks with various values of topological characteristics through controlling two input parameters. For this purpose, two parameter K and P are introduced, K for controlling the strength of preferential attachment and P for controlling clustering coefficient. It is found out on-line social network can be generated with the combinations of K(0~10) and P(0.3~0.5) or K=0 and P=0.9. Under these combinations of P and K small-world and scale-free properties are well represented. Node degree distribution follows power-law. Clustering coefficients are between 0.130 and 0.238, and average shortest path distance between 5.641 and 5.985. It is also found that on-line social network properties are maintained as network node size increases from 5,000 to 10,000.

키워드 : 온라인 소셜 네트워크, Small-World, Scale-Free, 네트워크 발생 모델, Power-Law 분포

Key words : On-line Social Network, Small-World, Scale-Free, Network Generation Model, Power-Law Distribution

Received 20 April 2020, **Revised** 1 May 2020, **Accepted** 16 May 2020

* **Corresponding Author** Kang-Won Lee(E-mail:kwlee@snut.ac.kr, Tel:+82-2-970-6476)
Professor, Department of Industrial Engineering, Seoul National University of Science and Technology, Seoul, 01811 Korea

Open Access http://doi.org/10.6109/jkiice.2020.24.7.914 print ISSN: 2234-4772 online ISSN: 2288-4165

ⓒThis is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/li-censes/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright Ⓒ The Korea Institute of Information and Communication Engineering.

Ⅰ. 서 론

인터넷과 이동 통신 기기(Mobile Communication Device)들이 진화해 나감에 따라 Facebook, Myspace 그리고 Flickr 등과 같은 온라인 소셜 네트워크는 개인 간 관계에 지대한 영향을 미치고 더 나아가 인간의 소셜 네트워크 자체를 재구성하기에 이르렀다. 주로 컨텐츠(Contents) 중심의 Web과 다르게 온라인 소셜 네트워크는 사용자 중심으로 이루어진다. 사용자들은 네트워크에 참여하여 그들의 Profile과 다양한 컨텐츠를 생산하고 그들이 관계 맺기를 원하는 다른 사용자들에게 연결 링크를 만든다. 이렇게 만들어진 소셜 네트워크는 사회적 관계(Social Relationship)를 유지하거나 다른 사용자들에 의해서 생산된 콘텐츠들을 공유하기 위한 기본적인 틀을 제공해 준다[1,2].

그런데 현재 광범위하게 이용되고 있는 온라인 소셜 네트워크의 효율성(Efficiency)을 분석한다거나 미래의 보다 진화된 온라인 소셜 네트워크를 설계, 또는 온라인 소셜 네트워크가 인터넷에 미치는 영향력을 분석하기 위해서는 무엇보다도 현재 사용 중인 온라인 소셜 네트워크의 위상적 구조와 특성(Topology Structures and Characteristics)에 대한 심층적인 이해가 필요하다[3,4,5]. 예로 온라인 소셜 네트워크의 구조에 대한 이해는 가장 영향력(Influential) 있거나 신뢰성(Trusted) 있는 사용자를 찾아내는 알고리즘 개발을 가능하게 한다[6]. 아울러 Worm-Virus 확산 방지 문제, 광고 프로모션(Promotion) 문제, Social-Aware Routing Protocol 설계 등 다양한 연구 분야에서 온라인 소셜 네트워크의 구조에 대한 이해는 필수적이다.

그런데 온라인 소셜 네트워크의 방대한 크기, 매우 복잡한 위상적 구조 그리고 보안(Security) 등의 문제로 온라인 소셜 네트워크의 위상적 구조를 심층 분석하고 이해한다는 것은 매우 어려운 일이다. 게다가 실제적인 소셜 네트워크에의 구조에 대한 정보를 광범위하게 채집해서 분석한다는 것은 간단한 일이 아니다. 그럼에도 불구하고 온라인 소셜 네트워크의 위상적 특성에 대한 일부 연구가 진행되어 왔다. Mislove[7] 등은 4개의 가장 인기 있는 온라인 소셜 네트워크(Flicker, YouTube, LiveJournal 그리고 Orkut)에 대한 대규모(11,300,000 사용자, 328,000,000 링크)실측 조사와 분석을 시도하였다. 분석 결과는 Web등 기존에 존재하는 여타 네트워크와 매우 다른 특성을 보이는 것으로 나타났는데 Power-Law, Small-World 그리고 Scale-Free 등의 성질이 나타나는 것을 확인하였다. Ahn[8] 등은 한국 온라인 소셜 네트워크인 Cyworld의 전체 네트워크와 Cyworld, Orkut 그리고 Myspace의 샘플 네트워크를 분석하였다. 온라인 소셜 네트워크의 주요 특징을 살펴보기 위하여 노드 차수 분포(Node Degree Distribution), 군집화 계수(Clustering Coefficient), 차수 상관관계(Degree Correlation) 그리고 평균 최단거리(Average Shortest Path Distance) 분석을 실시하였다. Sun[9]등은 중국 학생들의 온라인 소셜 네트워크를 실측하여 네트워크 위상 분석을 수행하였다. 노드 차수는 Power-Law 분포를 따르고 아울러 Small-World 성질도 갖고 있는 것으로 확인되었다.

한편 'Social-Aware Routing Protocol' 설계, 고장이나 Worm-Virus의 확산 문제, 그리고 광고 Promotion 확산 문제 등에 답하기 위해서 실제 소셜 네트워크상에서 이들 실험을 수행하고 분석가가 원하는 관련 측정 결과를 수집한다는 것은 간단한 일이 아닐뿐더러 법적인 문제까지 발생할 수 있다. 따라서 위에서 언급한 과거 연구자들이 밝혀낸 온라인 소셜 네트워크의 위상적 특징들을 재현 해낼 수 있는 인공(Artificial) 네트워크를 구축하고 그 위에서 관련 실험들을 수행하는 것이 합리적이고 합법적인 대안일 수밖에 없다. 이를 위해 최근 수년에 걸쳐서 온라인 소셜 네트워크 발생모델에 관한 연구들이 진행되었다. Rezvanian[10] 등은 소셜 네트워크를 위한 모델로 추계적(Stochastic) 그래프를 제시하였으며 Truong[11] 등은 소셜 네트워크 분석을 위한 그래프 모델들을 제시하였다. Bu[12] 등은 온라인 소셜 네트워크 발생 모델인 'Last Updating Evolution Model'을 제안하였다. 네트워크에 새롭게 참여하는 사용자는 다수개의 노드에 연결되는데 이때 노드 선정은 네트워크 전체가 아닌 일부 노드들만을 대상으로 'Preferential Attachment(PA)' 과정을 통해 이루어진다. 대상이 되는 노드들은 최근 'Updating Time'이 Threshold Θ 보다 작은 노드들로 구성된다. 그들은 Θ 값이 적절하게 선정되면 발생된 인위적인 망이 Small-World 성질을 보여 주고 노드차수가 Power-Law 분포를 따르는 것을 확인하였다.

기존의 온라인 소셜 네트워크 생성 모델 연구들은 그들이 생성한 네트워크가 온라인 소셜 네트워크가 갖고 있는 성질인 Small-World나 Scale-Free 성질을 갖고 있

는 것을 보여준다. 이를 위해 모델을 통해 생성된 네트워크의 노드 차수가 Power-Law 분포를 따르고 랜덤 네트워크에 비해 군집화 계수가 비교적 크고 노드 간 평균 최단 거리가 비교적 작다는 것을 보인다. 그런데 Mislove[8]의 연구에서 볼 수 있듯이 실제 온라인 소셜 네트워크들이 Small-World와 Scale-Free 성질을 갖고 있지만 그들의 위상 특성치들은 매우 다양한 값을 갖는 것으로 나타났다. 즉, 군집화 계수는 온라인 소셜 네트워크 종류에 따라 0.130에서 0.313까지의 값을 갖으며 평균 최단거리도 4.25에서 5.88, Power-Law 지수도 1.50에서 1.74까지 다양한 값들을 나타낸다. 따라서 온라인 소셜 네트워크 발생 모델도 기존의 단순히 Small-World 나 Scale-Free 성질을 표현하는 것에서 벗어나 사용자가 원하는 다양한 위상 특성치 값들을 적절하게 나타내 줄 수 있는 발생 모델이 필요하다.

따라서 본 연구의 목적은 실제적인 소셜 네트워크에 존재하는 Small-World와 Scale-Free 성질뿐만 아니라 네트워크의 특정 위상적 특성치(Topology Characteristics) 값들을 재생산 할 수 있는 인공적인 네트워크 발생 모델을 구축하는 것이다. 본 연구에서는 두개의 파라메터를 갖는 온라인 소셜 네트워크 생성 모델을 제안하고 이 파라메터들의 조절을 통해 원하는 다양한 위상 특성치들을 갖는 네트워크를 생성해 나가는 과정을 보였다. 이를 위해서 본 연구는 서론에 이어 2장에서 소셜 네트워크의 위상적 특성을 살펴보았다. 3장에서는 소셜 네트워크 발생 모델에 대해 다루었으며 4장에서는 본 연구에서 제안한 발생 모델을 통해 구축한 네트워크들을 분석해보고 5장에 최종 결론을 수록하였다.

Ⅱ. 온라인 소셜 네트워크의 위상적 특성

본장에서는 기존의 여러 연구에서 밝힌 온라인 소셜 네트워크가 갖는 위상적 특성에 대해서 살펴보았다.

2.1. Small-World 성질

최근 수년에 걸친 연구와 실측[7,8,9]에 따르면 온라인 소셜 네트워크는 Small-World 성질을 갖는 것으로 나타났다. Small-World 성질이란 일반적으로 온라인 소셜 네트워크의 막대한 크기에도 불구하고 네트워크 내 어떤 임의의 두 노드를 연결하는 상대적으로 짧은 경로가 존재한다는 것을 의미한다. 다음 <표 1>은 Mislove [7] 등이 실측을 통해 구한 여러 소셜 네트워크의 평균 최단거리를 나타낸다.

Table. 1 Average Shortest Path Distance of Social Network Obtained through Measurements

Network	Average Shortest Path Distance	Radius	Diameter
Flickr	5.67	13	27
LiveJournal	5.88	12	20
Orkut	4.25	6	9
Youtube	5.10	13	21

소셜 네트워크의 상대적으로 짧은 평균 최단거리는 네트워크의 노드 간 양방향 상호 연결성(Reciprocity)에 기인하는 면도 있지만 근본적으로 온라인 소셜 네트워크에 존재하는 다음의 두 가지 성질에 기인하는 것으로 보인다.

- 네트워크의 인접 노드 간 상호 연결성이 매우 높다. 이는 소셜 네트워크에서 '한 사용자의 친구들끼리는 서로 친구일 가능성이 상당히 높다'라는 사실로부터 확인할 수 있다. 이는 온라인 소셜 네트워크의 상당히 높은 군집화 계수로 나타난다. Ahn[8] 등은 Cyworld 전체 네트워크를 대상으로, MySpace 그리고 Orkut는 Snowball 샘플링을 통해 군집화 계수를 조사했는데 각각 0.16, 0.26, 그리고 0.31의 비교적 높은 값을 나타냈다. 그리고 Mislove[8] 등도 Flickr, LiveJournal, Orkut, YouTube의 군집화 계수들을 실측을 통해 확인하였는데 다음 <표 2>와 같이 각각 0.313, 0.330, 0.171 그리고 0.136으로 랜덤 네트워크에 비해 103배에서 105배에 이르는 것으로 나타났다.

Table. 2 Clustering Coefficient of Social Network Obtained through Measurements

Network	Clustering Coefficient	Ratio to Random Network
Flickr	0.31	47,200
LiveJournal	0.33	119,000
Orkut	0.17	7,240
Youtube	0.13	36,900

- 네트워크 내에 노드 차수가 매우 큰 허브 노드들이 존재한다. 이는 소셜 네트워크의 노드 차수가 Power-Law

분포를 따른다는 다수의 실측연구 결과 [7,8,9]로부터 확인할 수 있다. Mislove[7] 등은 Flickr, LiveJournal 그리고 YouTube 의 노드차수 분포가 Power-Law 분포를 따르는 것을 Kolmogorov-Smirnov 테스트를 통해 확인하였다.

2.2. Scale-Free 성질

네트워크의 특징을 나타내는 중요한 요소 중 하나는 노드가 갖고 있는 링크의 개수, 즉 노드 차수(Node Degree)다. 노드 차수 분포 P(k)는 임의의 노드가 k개의 링크를 갖는 확률을 나타낸다. 랜덤 네트워크의 경우 노드 차수는 대부분 비슷하고 평균 노드차수와 비슷한 값을 갖는다. 따라서 랜덤 네트워크의 경우 노드 차수 분포는 평균 노드 차수를 평균으로 갖는 Poisson 분포로 잘 나타낼 수 있다. 그런데 온라인 소셜 네트워크의 경우 실측을 통해 밝혀진 사실은 노드 차수 분포가 랜덤 네트워크와 다르게 Poisson 분포를 따르지 않고 Power-Law 분포를 따른다는 것이다. Power-Law 분포는 Poisson 분포와 다르게 꼬리(Tail)가 긴 전형적인 긴 꼬리 분포(Long Tail Distribution)다. 즉, 노드의 대부분은 노드 차수가 작으며 대신 노드 차수가 평균 노드 차수보다 매우 큰 슈퍼 노드가 다수 개 존재한다. 이런 Power-Law는 노드 차수가 어떤 특정 크기(Scale)와 무관하게(Free) 매우 큰 값을 가질 수 있기 때문에 Power-Law 분포를 따르는 네트워크를 Scale-Free 네트워크라고 부른다. 다음 <표 3>은 실측을[7] 통해 온라인 소셜 네트워크, Flickr, LiveJournal 그리고 Youtube의 노드차수가 Outdegree, Indegree 모두 Power-Law 분포를 따르는 것을 보이고 있다. 한편 Orkut는 노드 차수가 Power-Law 분포를 따르지 않는 것으로 나타났다.

Table. 3 Power-Law of On-Line Social Network

Net -work	Outdegree Power-Law Exponent	Power-Law Property	Indegree Power-Law Exponent	Power-Law Property
Flickr	1.74	O	1.78	O
Live -Journal	1.59	O	1.65	O
Orkut	1.50	X	1.50	X
You-Tube	1.63	O	1.99	O

네트워크의 노드 분포가 Power-Law 분포를 따르는지 확인하는 방법은 노드 개수와 빈도(frequency)를 Log-Log Scale로 작성하여 그것의 기울기가 대략 -1과 -3 사이의 직선식을 따르는지 보는 것 인데 주로 Kolmogorov-Smirnov 테스트를 수행한다.

Ⅲ. 소셜 네트워크 생성 모델

3.1. 기존 연구

현실의 대규모 망에 대한 상세한 위상(Topological) 정보 부재라든지 등의 이유로 Erdos와 Renyl의 Random 네트워크[13]가 최근 까지 네트워크의 구조 분석을 위해 유일하고 가장 광범위하게 사용되어온 모델이다. Small-World와 Scale-Free 성질의 발견은 이러한 성질들을 갖는 Small-world와 Scale-free 네트워크 개발을 위한 네트워크 생성 모델(Network Generation Model)에 대한 연구를 촉발시켰다. Small-World 네트워크 모델은 Watts 와 Strogatz[14]에 의해 처음으로 WS 모델이 제안되고 Scale-Free 네트워크 모델은 Barabasi와 Albert[15]에 의해 처음으로 BA 모델이 제안된 이후에 여러 다양한 모델들이[16,17,18] 개발되었다.

한편 네트워크 생성을 위해서는 그 용도에 따라 다양한 모델이 연구되어 왔다. 생화학 네트워크를 위한 모델, 온라인 소셜 네트워크 생성을 위한 모델 그리고 인터넷 ISP 도메인 네트워크 생성 모델 등 다양한 모델이 연구되어 왔다[11,12,19]. 그런데 이들 모델은 단편적으로 특정 네트워크를 생성해 줄 뿐 다양한 유형의 네트워크나 다양한 네트워크의 위상 측도값들을 생성하지는 못한다. 본 연구의 목적은 2장에서 살펴본 온라인 소셜 네트워크의 다양한 위상적 특성을 갖는 네트워크를 생성하는 것이다. 따라서 Small-World, Scale-Free 성질을 갖고 있으며 우리가 원하는 다양한 값의 군집화 계수, 노드 간 평균 최단거리, 그리고 다양한 종류의 노드 차수 분포 등을 생성해 줄 수 있는 네트워크 생성 모델이 필요하다.

위의 논의를 바탕으로 본 연구에서는 다음의 세 가지 요구 조건들을 만족하는 온라인 소셜 네트워크 생성 모델을 제안하였다.

- 생성 모델을 통해서 구축한 네트워크는 위에서 언급한 온라인 소셜 네트워크의 2가지 성질, 즉 Scale-Free, Small-World 성질을 갖고 있어야 한다.

- 온라인 소셜 네트워크는 2장의 온라인 소셜 네트워크의 위상적 특성에서 살펴 본 것처럼 Scale-Free나 Small-World 성질 중 하나만 갖고 있거나 아니면 두 성질 모두 다 갖고 있을 수 있다. 이를테면 2장에서와 같이 Flickr는 노드차수가 Power-Law 분포를 따르는 Scale-Free 네트워크다. 그리고 군집화 계수는 0.313로 매우 높고 노드 간 평균 최단 거리도 5.67로 비교적 짧은 Small-World 네트워크다. 반면에 Orkut는 노드차수가 Power-Law 분포를 따르지 않아 Scale-Free 네트워크는 아니지만 군집화 계수가 0.171로 높고 노드 간 평균 최단 거리가 4.25로 매우 짧은 Small-World 네트워크다. 따라서 본 소셜 네트워크 생성 모델은 모델의 파라메터를 적절하게 조절함으로서 Scale-Free나 Small-World 아니면 두 성질 모두 다 갖고 있는 네트워크를 생성할 수 있어야 한다.

- 일반적으로 소셜 네트워크는 다양한 값의 네트워크 위상 특성치 값들을 갖는다. 군집화 계수는 온라인 소셜 네트워크 종류에 따라 0.136에서 0.313까지의 값을 갖으며 평균 최단거리도 4.25에서 5.88, Power-Law 지수도 1.50에서 1.74까지 다양한 값들을 나타낸다. 따라서 온라인 소셜 네트워크 발생 모델도 기존의 단순히 Small-World나 Scale-Free 성질을 표현하는 것에서 벗어나 모델의 파라메터를 적절히 조절함으로서 사용자가 원하는 다양한 위상 특성치 값들을 나타내줄 수 있는 발생 모델이 필요하다.

3.2. 제안 모델 개요

본 모델은 기본적으로 Scale-Free 네트워크 구현을 위한 i) PA 과정, 그리고 인접 노드 간 연결 밀도를 높여서 Small-World 네트워크 구현을 위한 ii) Triad Formation(TF) 과정으로 이루어진다. 네트워크에 새롭게 들어온 노드의 차수는 지수 분포를 가정한다. PA 과정은 신규 노드가 기존 노드와 연결될 때 노드 차수가 큰 노드에 더 큰 확률을 부여하여 연결 노드를 선정한다. PA 과정 후 TF 과정은 확률 P로 발생하고 확률 1-P로는 PA 과정이 반복된다. TF 과정은 PA 과정을 통해 선정된 연결 노드의 친구들 중 하나와 신규 노드를 연결하는 과정인데 이는 소셜 네트워크에서 '친구의 친구와는 서로 친구일 확률이 높다'라는 사실을 구현하여 인접 노드 간 연결 밀도를 높이고자 한다. 이 두 개의 과정은 네트워크에 새롭게 들어오는 노드가 갖는 링크의 개수만큼 반복적으로 수행된다. 일례로 신규노드가 갖는 링크의 개수가 4라고 하고 P = 0.6이라고 하면 (그림 1)과 같은 절차를 따라 네트워크 생성이 진행된다. 한편 온라인 소셜 네트워크에서 Community Structure 성질은 TF 과정을 통해 어느 정도 구현된다.

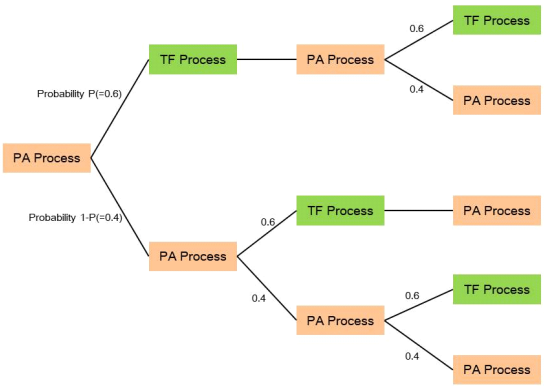

Fig. 1 Network Generation Process

3.3. PA 과정과 PF 과정

PA 과정은 기본적으로 Barabasi와 Albert의 BA[15] 모델을 토대로 한다. BA 모델은 PA를 기본으로 하여 Scale-Free 네트워크를 잘 구현할 수 있다. 다만 본 연구에서는 PA의 세기를 조절하기 위하여 신규노드 i가 기존 노드 j에 연결될 확률을 파라메터 K를 도입하여 다음 식 (1)과 같이 변형하였다.

$$P_j = \frac{d_j + \frac{K}{|s|}}{\sum_{i \in S} d_i + K} \tag{1}$$

dj는 노드 j의 차수 즉 링크의 개수를 나타내며 S는 현재 네트워크에 존재하는 노드들의 집합이고 ｜S｜는 노드들의 개수를 나타낸다. 식 (1)에서 K = 0이면 기존의 BA 모델이다. K가 0과 가까우면 PA로 인해 노드 차수가 Power-Law 분포를 따르고 슈퍼 허브 노드가 존재하게 된다. 이로 인해 노드 간 평균 최단거리도 감소하리라 기대된다. 반면에 K가 증가하면 PA 세기가 약해져 노드 차수가 Power-Law 분포에서 멀어지게 되고 슈퍼 허브 노드도 존재하지 않게 돼 노드 간 평균 최단거리의 증가가 예상된다. K가 무한대로 커지면 랜덤 네트워크화 된다. 즉, K를 조절함으로서 Scale-Free 네트워크를

생성할 수도 있고 못할 수도 있다. 반면에 군집화 계수는 K 값과는 큰 관계는 없어 보여서 K의 조절을 통해 Small-World 성질을 표현할 수는 없다. 다음 <표 4>는 K 값에 따른 위상 특성치의 예상 변화를 요약 정리 하였다. 이에 대한 구체적인 검증은 4장 네트워크 생성 및 분석의 "K와 P의 변화에 따른 네트워크의 위상 특성치 값들의 변화"에서 실증적으로 상세하게 논하였다.

Table. 4 Network Topological Characteristics according to K

	K (0 ↔ ∞)	
Scale-Free Property	Strong	Weak
Small-World Property	No Effect	
Power-Low Property	Strong	Weak
Average Shortest Path Distance	Small	Large
Clustering Coefficient	Little Effect	

TF 과정은 PA 과정을 통해 연결된 기존 노드의 친구들 중 하나와 연결하는 과정이다. 이때 기존 노드의 친구들은 임의로 선정된다. 다음 (그림 2)에서처럼 신규 노드 v가 PA 과정을 거쳐 기존 노드 w에 연결 되었다고 하자. 이때 w는 3명의 친구 a, b 그리고 c를 갖고 있는데 이중 임의로 친구 b가 선정되는 TF 과정을 보여준다.

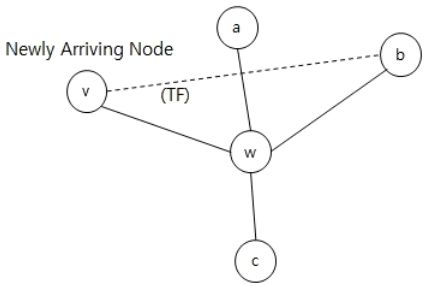

Fig. 2 TF Process

TF 과정은 PA 과정 후 신규 노드에 연결 링크수가 처음에 신규 노드에 부여된 링크 개수보다 적을 경우 확률 P로 수행된다. 이때 P를 1에 가깝게 크게 정하면 TF 과정이 자주 수행되므로 인접 노드 간 연결 밀도가 높아져 군집화 계수의 증가가 기대된다. 반면에 PA 과정 수행 빈도가 줄어들게 되므로 노드 차수가 Power-Law 분포를 따를 가능성이 줄어든다. 즉, P 값이 증가하면 할수록 Small-World 성질은 강하게 나타나지만 Scale-Free 성질은 점점 감소하게 된다. 반대로 P값이 0에 가까워지면 TF 과정이 거의 수행되지 않기 때문에 군집화 계수 값이 감소하리라 기대된다. 대신에 PA 과정은 더 많이 수행되므로 노드 차수가 Power-Law 분포를 따를 가능성이 늘어난다. 즉, P 값이 감소하면 할수록 Small-World 성질은 점점 사라지지만 Scale-Free 성질은 점점 강하게 나타난다. 그리고 반대로 P 값이 1에 근접할수록 Small-World 성질은 점점 증가하지만 Scale-Free 성질은 점점 약하게 나타난다.

한편 노드 간 평균 최단거리는 P 값의 변화에 따른 영향이 분명하지 않다. P 값이 1에 근접하면 군집화 계수가 늘어나서 노드 간 평균 최단거리가 감소할 수도 있지만 반면에 노드 차수가 Power-Law 분포를 따를 가능성이 줄어들어 슈퍼 허브 노드의 존재로 인한 노드 간 평균 최단 거리의 감소 현상이 발생하지 않는다. 반대로 P 값이 0에 근접하면 군집화 계수가 작아져서 노드 간 평균 최단거리가 증가할 수도 있지만 반면에 노드 차수가 Power-Law 분포를 따를 가능성이 커져서 슈퍼 허브 노드의 존재로 인한 노드 간 평균 최단 거리의 감소 현상이 발생한다. 이와 같이 P 값에 따른 노드 간 평균 최단거리의 변화는 증가요인과 감소 요인이 동시에 나타나기 때문에 현 시점에서는 P 값의 노드 간 평균 최단거리에 대한 영향력이 불분명하다. 다음 <표 5>는 P 값에 따른 네트워크 위상 특성치의 예상 변화를 요약 정리 하였

Table. 5 Network Topological Characteristics according to P

	P (0 ↔ 1)	
Scale-Free Property	Strong	Weak
Small-World Property	Weak	Strong
Power-Low Property	Strong	Weak
Average Shortest Path Distance	Not Certain (will be discussed in Chapter 4.)	
Clustering Coefficient	Weak	Strong

다. P 값의 노드 간 평균 최단거리에 대한 영향력과 아울러 다른 구체적인 논의는 다음 4장 네트워크 생성 및 분석의 “K와 P의 변화에 따른 네트워크의 위상 특성치 값들의 변화”에서 실증적으로 상세하게 할 예정이다.

Ⅳ. 네트워크 생성 및 위상 분석

본 장에서는 3장에서 제안한 네트워크 생성 모델을 이용하여 두 개의 파라메터 K와 P를 변화시켜 나가면서 다양한 네트워크를 생성한 후 이들의 위상적 특성을 분석하였다.

- 총 노드 수 5,000개를 갖는 네트워크를 생성하였다.
- 신규로 들어오는 노드의 링크 개수는 평균 3개의 지수 분포로 가정하였다. 소수점은 반올림하여 사용하였는데 최소 1개의 링크는 갖도록 하였다.
- 특정 K와 P의 조합에서 각 위상 특성치 값들은 네트워크 10개의 평균 특성치 값을 사용하였다. K 값은 0, 10, 100, 1,000 네 개의 값을 P 값은 0, 0.3. 0.5 그리고 0.9 네 가지로 하였다.

4.1. 네트워크 생성 및 결과

K와 P의 조합 16개의 각각 경우에 따라 노드 수 5000개를 갖는 인공 네트워크를 10개 생성하였다. 그리고 특정 K와 P의 조합에 따른 위상 특성치를 10개 인공 네트워크 위상 특성치들의 평균으로 다음 <표 6>에 나타냈다.

4.2. 분석

본 연구의 분석은 다음의 세 가지 방향으로 수행하였다.

- 먼저 K와 P의 변화에 따른 네트워크의 위상 특성치 값들의 변화를 살펴보고 이들 조합에 따라 어떤 특성을 지닌 네트워크를 생성할 수 있는지 조사하였다.
- 이를 토대로 본 연구의 목적인 소셜 네트워크를 생성하는 방법을 제안하였다.
- 네트워크 크기(노드 크기)에 따른 위상 특성치 값들의 변화를 조사하였다.

Table. 6 Average Network topological Characteristics according to Combination of K and P

		P = 0.1	P = 0.3	P = 0.5	P= 0.9
K = 0	Average Node Degree	2.936	2.942	2.903	2.874
	Maximum Node Degree	300	264	246	206
	Node Degree Distribution(Follow Power-Law:1, otherwise: 0)	1	1	1	1
	Average Clustering Coefficient	0.052	0.131	0.169	0.238
	Average Shortest Path Length	5.478	5.641	5.759	5.985
	Network Diameter	12	14	14	15
K = 10	Average Node Degree	2.966	2.895	2.917	2.939
	Maximum Node Degree	252	244	204	158
	Node Degree Distribution(Follow Power-Law:1, otherwise: 0)	1	1	1	0
	Average Clustering Coefficient	0.051	0.130	0.169	0.239
	Average Shortest Path Length	5.513	5.799	5.878	6.058
	Network Diameter	13	14	14	16
K = 100	Average Node Degree	2.907	2.910	2.892	2.906
	Maximum Node Degree	176	166	148	136
	Node Degree Distribution(Follow Power-Law:1, otherwise: 0)	0	0	0	0
	Average Clustering Coefficient	0.052	0.122	0.175	0.242
	Average Shortest Path Length	5.740	5.887	6.061	6.110
	Network Diameter	14	15	15	16
K = 1,000	Average Node Degree	2.944	2.894	2.899	2.927
	Maximum Node Degree	108	92	83	76
	Node Degree Distribution(Follow Power-Law:1, otherwise: 0)	0	0	0	0
	Average Clustering Coefficient	0.051	0.112	0.168	0.236
	Average Shortest Path Length	6.121	6.269	6.366	6.465
	Network Diameter	15	15	16	16

1) K의 변화에 따른 네트워크의 위상 특성치 값들의 변화

- K에 따른 Power-Law 성질과 Scale-Free: <표 6>을 살펴보면 본 연구에서 사용한 네트워크 생성 모델은 K = 0 일 때는 P값에 무관하게 항상 Power-Law 분포를 따르며 K = 10일 때는 P = 0.9 일 때를 제외하고는 모두 Power-Law 분포를 따른다. P = 0.9일 때 Power-Law를 따르지 않는 이유는 TF 과정 수행 확률이 높아져 PA 과정이 상대적으로 적게 수행되기 때문이다. 그 외 K = 100이나 1,000일 때는 Power-Law 분포를 따르지 않는다. 2장의 예상대로 K 값의 조절에 따라 슈퍼 허브가 존재하여 노드 차수 분포가 Power-Law를 따르는 Scale-Free 네트워크를 구축할 수도 있고 못할 수도 있다.

- K 값과 군집화 계수: 노드 밀도와는 상관이 없기 때문에 군집화 계수 값은 예상대로 K 값과 무관하게 나타났다. 예로 P가 0.1일 때 군집화 계수는 다음 (그림 3)에서 보듯이 거의 일정함을 볼 수 있다.

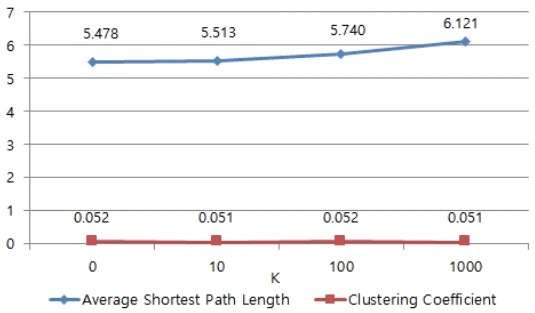

Fig. 3 Clustering Coefficient and Average Shortest Path Distance according to K(P = 0.1)

반면에 평균 최단거리는 (그림 3)에서 보듯이 K 값이 증가할수록 조금씩 증가하는 양상을 보인다. K = 0 일 때 평균 최단거리가 5.478로 가장 낮게 나타났는데 이는 K = 0 일 때는 K = 10이나 100 일 때에 비해 보다 강력한 슈퍼 허브 노드가 존재해 노드 간의 이동을 간편하게 해주는데 기인한 것으로 보인다. 한편 네트워크 Diameter도 K = 0 일 때 12로 100이나 1,000일 때의 14나 15에 비해 작음을 볼 수 있다.

2) P의 변화에 따른 네트워크의 위상 특성치 값들의 변화

- P와 군집화 계수: <표 6>을 살펴보면 본 연구에서 사용한 네트워크 생성 모델은 P가 증가할수록 노드 밀도가 증가하기 때문에 평균 군집화 계수도 늘어난다. 다음 (그림 4)는 K = 10일 때 P값의 변화에 따른 군집화 계수의 변화를 나타낸다. P = 0.1 일 때 0.051에서 0.130(P = 0.3), 0.169(P = 0.5) 그리고 0.239(P = 0.9)로 증가함을 볼 수 있다.

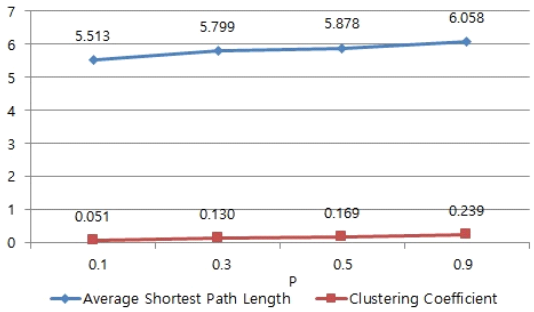

Fig. 4 Clustering Coefficient and Average Shortest Path Distance according to P(K = 10)

- P와 평균 최단거리: P가 증가하면 노드 밀도가 증가하기 때문에 당연히 평균 최단 거리도 감소해야한다. 그런데 (그림 4)에서 보면 반대로 P가 증가할수록 평균 최단거리도 증가함을 볼 수 있다. 3.절의 네트워크 생성 모델은 PA 과정 후 확률 P로 Triad Formation과정을 수행하는데 반복 과정 횟수가 신규 진입 노드가 갖는 링크 수에 도달할 때까지 반복적으로 수행한다. 따라서 P가 증가할수록 PA 과정 수행 횟수는 줄어들게 되고 이러면 슈퍼 허브 노드들의 출현 확률이 낮아지게 되므로 평균 최단거리는 늘어나게 된다. P가 0일 때는 Triad Formation 과정을 수행하지 않고 신규 진입 노드가 갖는 링크 수만큼 PA 과정이 수행되므로 강력한 슈퍼 허브 노드들이 출현할 가능성이 높아지게 되므로 평균 최단거리도 가장 작은 값을 갖는다. 네트워크 Diameter도 P가 증가할수록 다소 증가하는 양상을 보인다. 이를 통해 네트워크의 평균 최단거리에 가장 영향을 미치는 요소는 군집화 계수가 아니라 슈퍼 허브 노드의 존재를 가능하게 하는 Power-law 성질이라는 것을 확인할 수 있었다.

3) 소셜 네트워크 생성

2장에서 살펴본 소셜 네트워크가 갖고 있는 위상적 성질은 먼저 짧은 평균 최단거리(4.25 ~ 5.88)와 높은 군집화 계수(0.130~0.330)를 갖는 Small-World 성질이다. 아울러 온라인 소셜 네트워크는 슈퍼 허브노드의 존재로 인해 대부분 노드 차수 분포가 Power-Law 분포를 따르는 것으로 나타나 Scale-Free 성질도 내포하고 있다.

평균 최단 거리가 얼마 이하고 군집화 계수는 얼마 이상이어야 Small-World 성질인지에 대한 객관적인 기준은 존재하지 않는다. Small-World 성질은 이름이 말하듯이 네트워크의 구성 노드들 간의 평균 최단 거리가 네트워크의 크기와 무관하게 작다는 것이다. 사회 과학자 Milgram[20]에 의해 제안된 'Six Degree of Separation'은 '네트워크에서 임의의 한 노드는 평균적으로 6개의 노드를 거치면 모든 노드에 도착할 수 있다'라는 것을 의미하고 일반적으로 평균 최단 거리 6이 Small-World를 나타내는 지표로 널리 사용되고 있다. 본 연구에서도 평균 최단거리 기준은 6으로 하고 군집화 계수는 2장의 실측 자료를 토대로 하였다. 즉, 평균 최단거리가 6 이하고 군집화 계수가 0.130 이상이 되는 것을 Small-World 성질로 정하였다. 한편 Scale-Free 성질은 Kolmogorov-Smirnov 테스트를 통해 노드 차수 분포가 Power-Law를 따르는지 여부를 기준으로 삼았다.

어떤 K와 P의 조합이 Small-World 성질과 Scale-Free 성질을 갖는 네트워크를 생성하는지 다음 <표 7>과 <표 8>에 각각 나타냈다. 음영 처리된 부분이 Small-World 성질이나 Scale-Free 성질을 갖는 네트워크를 나타낸다.

Table. 7 Combination of K and P for Small-World Property

	P = 0.1	P= 0.3	P = 0.5	P = 0.9
K = 0				
K = 10				
K = 100				
K = 1,000				

Table. 8 Combination of K and P for Scale-Free Property

	P = 0.1	P= 0.3	P = 0.5	P = 0.9
K = 0				
K = 10				
K = 100				
K = 1,000				

위의 <표 7>과 <표 8>로부터 다음을 확인할 수 있다.

- 먼저 본 연구의 모델을 통해 생성된 네트워크는 Small-World 성질을 갖고 있으면 Scale-Free 성질도 동시에 갖고 있는 것으로 나타났다. 반면에 Scale-Free 성질을 갖는 네트워크 중에는 Small-World 성질을 갖고 있지 않은 네트워크들이 존재할 수 있다.
- <표 7>과 <표 8>에서 볼 수 있듯이 K = 0과 10 그리고 P가 0.3과 0.5 사이의 조합이나 K = 0과 P = 0.9의 값을 사용하여 생성한 네트워크는 노드 차수 분포가 모두 Power-Law 분포를 따라서 Scale-Free 성질을 잘 나타내고 있다. 군집화 계수 0.130 ~ 0.238, 평균 최단거리 5.641 ~ 5.985로 앞에서 언급한 기준치 안에 속하므로 Scale-Free 성질도 나타낸다. 따라서 K = 0과 10 그리고 P가 0.3과 0.5 사이의 조합이나 K = 0과 P = 0.9의 값을 이용하면 소셜 네트워크를 생성할 수 있다.
- 이제 Small-World 성질과 Scale-Free 성질을 동시에 보여주는 K와 P의 값에 따라 위상 특성치가 어떻게 변하는지 다음 <표 9>에서 4가지 경우를 살펴보았다.

앞에서 논의한 대로 인공 네트워크 발생을 원하는 사용자는 K = 0과 10 그리고 P가 0.3과 0.5 사이의 조합이나 K = 0과 P = 0.9의 값을 이용하면 소셜 네트워크를 생성할 수 있다. 그런데 생성된 인공 소셜 네트워크에서 사용자가 원하는 위상 특성치는 위의 <표 9>를 바탕으로 어느 정도 구현 가능하다. 예로 군집화 계수는 조금 낮더라도 Power-Law 성질이 강하게 나타나서 평균 최단 거리가 짧은 소셜 네트워크를 원한다면 K = 0과 P = 0.3 부근의 조합이 좋다. 그리고 Power-Law 성질은 다소 약하게 나타나지만 군집화 계수가 높은 소셜 네트워크를 원하면 K = 0과 P = 0.9 부근의 조합이 좋다.

Table. 9 Topological Characteristics according to Combination of K and P

K = 0, P = 0.9	- Power-Law Property(4) Average Shortest Path Distance 5.985 Clustering Coefficient 0.238
K = 0, P = 0.3	- Highest Power-Law Property (1) Average Shortest Path Distance5.641 Clustering Coefficient 0.131
K = 10, P = 0,3	- Power-Law Property (2) Average Shortest Path Distance 5.799 Clustering Coefficient 0.130
K = 10, P = 0,5	- Power-Law Property (3) Average Shortest Path Distance 5.878 Clustering Coefficient 0.169

(The number in the parenthesis represents the rank of strength of power-law property)

- 그런데 본 연구에서 제안한 모델을 통하여 생성할 수 있는 소셜 네트워크는 군집화 계수가 증가하면 평균 최단거리도 따라서 증가하고 평균 최단거리를 낮추면 군집화 계수도 낮아진다. 이유는 K와 P의 변화에 따른 네트워크의 위상 특성치 값들의 변화에서 설명했듯이 파라메터 P값의 변화에 따라 군집화 계수와 평균 최단거리가 동시에 증가하거나 감소하기 때문이다. 따라서 평균 최단거리는 작으면서 군집화 계수는 높거나 평균 최단거리는 비교적 크면서 군집화계수가 큰 소셜 네트워크는 생성할 수 없다. 다시 말하면 군집화 계수와 평균 최단거리를 각각 독립적으로 조정할 수 없다는 것이 본 연구에서 제안한 모델의 단점일 수 있다.

4) 네트워크 크기에 따른 변화

다음은 네트워크의 크기(노드 개수)가 증가할 때 보이는 성질에 대하여 살펴보았다. 이를 위해 소셜 네트워크를 생성하는 조합인 K = 0과 P = 0.5의 경우 노드 크기를 현 5,000개에서 10,000개까지 늘려가면서 군집화 계수와 평균 최단거리의 추이를 살펴보았다. (그림 5)에서 보듯이 먼저 평균 최단거리는 노드 개수가 5,000개 일때 5.759인데 노드 개수를 2배인 10,000개로 늘려도 평균 최단거리는 5.906으로 변화가 크지 않다. 군집화 계수도 0.169에서 0.177로 노드 개수와 무관하게 거의 비슷하다. 한편 Power-Law 성질은 노드 크기를 10,000개로 늘려도 계속 유지되는 것으로 나타났다. 즉, 노드 개수가 늘어나도 Small-World 성질은 그대로 유지되고 Power-Law 성질도 노드 개수와 무관하게 존재해서 본 연구에서 제안한 모델은 노드 개수와는 무관하게 소셜 네트워크를 생성해주는 것으로 나타났다.

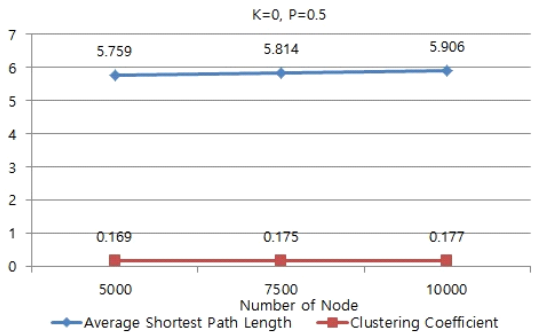

Fig. 5 Clustering Coefficient and Average Shortest Path Distance according to Number of Node

Ⅴ. 결 론

본 연구에서는 소셜 네트워크를 생성할 수 있는 인공적인 네트워크 발생 모델을 제안했다. 모델은 소셜 네트워크의 특징인 Small-World 성질과 Scale-Free 성질을 효율적으로 나타낼 수 있도록 구축하였다. 먼저 Scale-Free 성질을 나타내기 위하여 PA 과정을 사용하였으며 PA의 세기를 조정 할 수 있도록 파라메터 K를 도입하였다. Small-World 성질을 나타내기 위해서 군집화 계수를 적절하게 조정할 수 있도록 파라메터 P를 도입하였다.

본 연구에서 제안 한 모델을 이용하여 네트워크들을 생성한 후 분석한 결과 Power-Law 성질은 K값에 따라 조절 가능한 것으로 나타났다. K가 0에 수렴 할수록 노드 차수의 Power-Law 성질이 강하게 나타나며 증가하면 Power-Law 성질이 약해진다. 군집화 계수는 P값을 이용하면 조정 가능한 것으로 나타났으며 P가 1에 근접할수록 군집화 계수는 증가한다. 평균 최단 거리는 K와 P값에 따라 변해나가는데 K가 큰 값을 갖을수록 그리고 P가 1에 수렴할수록 증가하며 그 반대의 경우는 감소하는 것으로 나타났다.

K = 0과 10 그리고 P가 0.3과 0.5 사이 혹은 K = 0과 P = 0.9의 조합을 이용하면 인공 소셜 네트워크를 생성할 수 있다. 노드 차수 분포가 모두 Power-Law 분포를 따라서 Scale-Free 성질을 잘 나타내고 있다. 군집화 계수 0.130 ~ 0.238, 평균 최단거리 5.641 ~ 5.985로 앞에서 언급한 기준치 안에 속하므로 Scale-Free 성질도 나타낸다. 아울러 K와 P의 조합에 따른 Power-Law 성질의 세기, 군집화 계수, 평균 최단 거리를 나타내는 표를 만들어서 사용자가 원하는 위성 특성 값은 갖는 인공 소셜 네트워크를 생성하도록 하였다. 또한 네트워크의 크기를 노드 5,000개에서 10,000개로 증가시켜도 소셜 네트워크 성질을 유지하는 것으로 나타났다.

본 연구에서 제안한 소셜 네트워크 생성 모델은 군집화 계수와 평균 최단 거리를 독립적으로 제어할 수 없다는 단점이 있다. 즉, 군집화 계수를 고정 시키고 평균 최단거리를 증감 시키거나 아니면 평균 최단거리를 고정 시키고 군집화 계수를 증감 시킬 수 없다는 점이다. 본 모델은 군집화 계수가 증가하면 평균 최단거리도 늘어나고 반대로 감소하면 줄어든다. 네트워크의 위상 특성치를 독립적으로 제어할 수 있는 보다 일반화된 소셜 네트워크 생성 모델은 추후 과제로 남긴다.

ACKNOWLEDGEMENT

This research was supported by Seoul National University of Science and Technology Research funds.

REFERENCES

[1] B. Jang and J. Yoon, “Characteristics analysis of data from news and social network services,” *IEEE Access*, vol.6, pp.18061-18073, 2018.

[2] V. Amati, A. Lomi, and A.Mira, “Social Network Model,” *Annual Review of Statistics and its Application*, pp.1-29, Mar. 2018.

[3] S. Tabassum, F. S. F. Pereira, S. Fernandes, and J. Gama, “Social Network Analysis: An Overview,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 58, pp.1- 30, Apr. 2018.

[4] A. V. Proskurnikov and R. Tempo, “A Tutorial on Modeling and Analysis of Dynamic Social Networks. Part 1,” *Annual Reviews in Control*, pp.1-21, May 2017.

[5] K. W. Lee and J. S. Lee, “Power-Law of Node Degree and Information Diffusion Process,” *The Journal of Korean Institute of Communication and Information Sciences*, vol. 44, no. 10, pp.1866-1877, 2019.

[6] J. Kim and M. Hastak, “Social network analysis: Characteristics of online social networks after a disaster,” *Information Journal of Information Magement*, vol.38, pp.86-96, 2018.

[7] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and analysis of online social networks,” *Internet Measurement Conference*, 2007.

[8] Y. Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, “Analysis of topological characteristics of huge online social networking services,” *WWW, Semantic Web and Web 2.0*, 2007.

[9] D. Sun, J. Wu, S. Zheng, B. Hu, and K. M. Carley, “Topological Analysis and measurements of an online chinese student social network,” *Complex, Part 1*, pp.737-748, 2009.

[10] A. Rezvanian and M. R. Meybody, “Stochastic graph as a model for social networks,” *Computers in Human Behavior*, vol.64, pp.621-640, 2016.

[11] Q. D. Truong, Q. B. Truong, and T. Tkaki, “Graph method for social network analysis,” *International Conference on Nature of computation and Communication*, pp.276-286, October 2016.

[12] Z. Bu, Z. Xia, J. Wang, and C. Zhang, “A last updating evolution model for online social networks,” *Physica A*, 392, pp.2240-2247, 2013.

[13] P. Erdos and A, Renyi, “On the evolution of random graph,” *Public Mathmatical Institute Hungary Academic Science*, vol.5, pp.17-60, 1959.

[14] D. J. Watts and S.H. Strogatz, "Collective dynamics of small-world," *Nature, 393*, pp.440-442,1998.

[15] A. Barabasi and R. Albert, "Emergence of Scaling in Random Networks," *Science 286*, pp.509-512, 1999.

[16] R. D. Caux, C. Smith, D. Kniveton, R. Black, and A. Philippides, “Dynamic, small-world, social network generation through local agent interactions,” *Complexity*, vol. 19, no. 6, pp. 44-53, 2014.

[17] K. W. Lee, H. K. Uhm and H. J. Choe, “Tunable network generation model for small-world and scale-free network,” *The Journal of Korean Institute of Communication and Information Sciences*, vol.42, no.07, pp.1392-1401, 2017.

[18] M. Q. Pasta and Z. Jan. (2014, October). Tunable and growing network generation model with community structures. [Online] arXiv:1310.8396v2[csS1].

[19] K. W. Lee, S. W. Moon, B. H. Kim and S. Y. Jun, “Generation method of Korea Internet AS network using BA model,” *The Journal of Korean Institute of Communication and Information Sciences*, vol.43, no.06, pp.986-994, 2018.

[20] S. Milgram, “The small world problem,” *Psychology Today*, vol.2, pp.60-67, 1967.

이강원(Kang-Won Lee)

1980년 서울대학교 산업공학과 학사
1982년 서울대학교 산업공학과 석사
1985년 Kansas State University 산업공학과 박사
1989년~현재 서울과학기술대학교
글로벌융합산업공학과 교수
※관심분야 : 정보통신, 품질 및 신뢰성, O.R., 차세대 이동통신, RFID