

연구논문

네트워크 생성 모형의 설명 구조: 바라바시-알버트 선호적 연결 모형을 중심으로*

손운규**

네트워크 이론이 사회과학을 비롯한 다양한 분과학의 보편적인 분석틀이 되고, 대표적인 학제적 분야로 성장하는 데에는 2000년을 전후로 발표된 네트워크 형성 모형에 관한 기념비적 논문들이 핵심적인 역할을 했다고 할 수 있다. 네트워크 생성 모형 대부분은 자연과학, 특히 물리학의 전형적인 이론 정초 형식을 따르고 있는데, 이러한 접근은 정성적 개념 간의 관계를 나타내는 사회 이론과 조작화를 통한 경험 검정에 익숙한 사회과학자들에게는 생소할 수 있다. 이 논문에서는 현대 네트워크 과학의 기원이 된 네트워크 생성 모형의 설명 구조를 과학 철학의 표준적인 관점에서 분석하고, 사회과학 경험 연구의 일반적인 설명 형식과 비교한다. 이를 바탕으로 사회 자료를 이용하는 분과학 간의 상호 이해를 모색하고, 사회과학 분야가 학제적 연구 경향에 기여할 수 있는 바를 논의한다.

주제어: 네트워크 생성 모형, 법칙 연역적 설명, 인과적 메커니즘, 자연과학, 사회과학

* 이 연구는 서울대학교 신입교수 연구정착금으로 지원되는 연구비에 의하여 수행되었음.

** 서울대학교 사회학과 조교수(ysohn@snu.ac.kr).

I. 문제의 제기

일상의 사례에서 학문적 탐구의 대상에 이르기까지, 설명을 통한 이해는 인간 지성의 원초적 욕구이자 문명의 근간이 되는 활동이라고 할 수 있다. 상식만이 필요한 설명에서 고도의 전문성을 요구하는 설명에 이르기까지 타당한 설명의 보편적인 구성 요건은 인식론과 과학 철학의 근본 주제이다. 다양한 상황에서의 설명이 공유하는 근본 구조가 존재하는 한편으로, 그 연구 대상을 오랫동안 공유해 온 분과학(discipline) 특유의 관행 혹은 문화가 사회적으로 형성되고 시대적으로 변화해 왔음도 사실이다.

사회과학의 경우, 연구 논문은 사회 현상의 보편적인 원인에 대한 진술인 이론을 특수한 사례 집합에 대한 분석을 통해서 검정하는 형식을 띤다. 정성적, 정량적 접근을 막론하고, 경험 사회과학자들은 추상적으로 정의된 개념들을 양화 혹은 측정 가능한 형태로 조작화한(operationalize) 뒤에 그들 사이의 관계에 대한 통계적 검정을 비롯한 일련의 논리적 절차를 거쳐서 이론 가설을 확인하는 데에 익숙하다. 이는 사회과학 분과학 일반의 표준적인 논증 형식으로 자리잡았다.

앞서 언급한 바와 같이, 사회과학적 전통 밖에 있는 분야의 연구 형식은 이러한 설명 방식과 공통점과 차이점을 모두 갖고 있다. 과학 철학 분야에서 타당한 설명의 형식에 대한 연구는 주로 자연과학, 특히 물리학의 사례를 바탕으로 연구되었고, 생물학과 심리학을 비롯한 특정 분과학 사례로 확장되고 있다(Okasha 2002). 사회과학의 설명 단위, 대상, 방법의 특수성을 논한 저서들은 기본적으로 자연과학에 대한 철학적 논의를 바탕으로 하고 있으며, 그 공통점과 차이점에 대한 정성적 분석을 수행했다(Elster 1989; Little 1991).

과학철학자가 아닌 경험 연구자가 본인이 속하지 않은 분야의 논증 방식을 이해할 필요에 대해 근본적인 의문을 제기할 수도 있다. 이는 그동안 대부분의 연구자들이 특정 분과학의 연구 문헌을 소화하고 이와 대동소이한 형식의 연구를 수행하는 것으로 충분했기 때문일 것이다. 하지만 오랫동안 독립적으로, 혹은 고립되어 발전한 분과학적 전통에 비겨, 현대의 분과학은 통계학의 기술을 비롯한 방법론적 도구를 서로 신속하게 공유하고 있으며, 사회 자료를 대상으로 하는 연구들을 포함해, 그 연구 대상 또한 공유하는 사례가 늘어나고 있다(Lazer et al. 2009). 다시 말해,

분과학의 벽을 넘지 않고는 유의미한 연구를 생산하기 힘든 시대가 도래하고 있다.¹⁾

그 대표적인 분야가 네트워크 과학이다. 네트워크 과학이 폭발적인 성장을 맞이한 데에 수리 통계적 방법으로 네트워크의 분포적 특성을 알아낸 2000년 전후의 기념비적 이론 연구들이 있었다는 주장을 부정할 학자는 많지 않을 것이다. 이로부터 출발한 열기와 관심은 오늘날 네트워크 자료에 대한 인과적 추론 방법, 그래프 신경 네트워크(graph neural network)와 같은 주제로까지 이어지면서, 네트워크는 과학 방법론 전반의 선두적인 개념 틀로 지속적으로 이용되고 있다. 사회과학, 자연과학, 공학, 인문학 등 학문 전 분야에 지대한 영향을 미치고 있는 네트워크 과학의 뿌리가 수학의 그래프 이론에서 영감을 받은 인류학, 사회학, 심리학의 고전적 연구들에 있다는 사실은, 그 태생부터 이미 학제적(interdisciplinary)임을 드러낸다고도 할 수 있다.

이 논문에서는 사회과학 분야에서도 매우 흔하게 인용되는 동시에, 네트워크 과학 분야에서 가장 많이 인용된 논문 중 하나²⁾인 Barabasi & Albert(1999)를 중심으로 네트워크 생성 모형 연구의 논증 방식을 분석하고, 그 특성과 한계에 대해서 논한다. 뒤이어 제안된 많은 네트워크 생성 모형과 마찬가지로 통계물리학자들에 의해 제안된 바라바시-알버트(Barabasi-Albert: BA) 모형의 체계적인 이해를 위해서, 설명에 대한 과학 철학에서의 표준적인 견해를 소개하고, 네트워크 생성 모형 연구가 물리학 전통의 설명 방식과 어떻게 맞닿아 있는지 보인다.

이를 통한 일차적인 목표는 사회과학자들에게 생소할 수 있는 네트워크 생성 모형 설명 구조의 특성을 사회과학 연구 일반의 설명 구조와의 비교를 통해 상세히 설명하는 것이다. 물리학 이론의 연역 체계에 강한 영향을 받은 미시경제학 이론을

1) 계산사회과학(computational social science)이라는 학제적 사회현상 연구 패러다임에 대한 선언적 성격의 논문인 Lazer et al.(2009)의 저자 15명 각각의 학문 분야는 사회학, 정치학, 경제학, 경영학 등 사회과학 분야는 물론, 전산학과 인공지능 분야로, 이후 전개될 학제성의 방향을 잘 보여준다고 할 수 있다. 이 논문에서 학제적 연구 대상의 주요 예시로 제시되는 것이 온라인 사회 네트워크를 비롯한 사회 네트워크 연구이다.

2) 2023년 현재 구글 스칼라(Google Scholar)를 기준으로 네트워크 과학 분야에서 가장 많이 인용된 논문은 일종의 네트워크 교란(perturbation) 모형이라고 할 수 있는 좁은 세상(small world) 모형을 제안한 Watts & Strogatz(1998)이며, Barabasi & Albert(1999)는 두 번째로 많이 인용되었다. 바라바시의 연구는 그가 저술한 대중서인 《링크》(Linked)의 성공에 더불어 사회과학자를 비롯한 비자연과학 배경의 학자들과 독자들에게 복잡계 네트워크(complex network) 연구의 대명사처럼 널리 알려졌다. 특히, 이 모형은 통계물리학자들에 의해서 제안된 모형 중 가장 성공적인 모형으로, 좁은 세상 모형을 비롯하여 지수적 무작위 네트워크 모형 등, 타분야의 학자들에 의해서 제안되거나 발전된 모형들에 비해, 물리학 전통과 그 접점을 찾는 본 논문의 논지에서 중요성이 크다고 할 수 있다. 높은 인지도를 갖는 다른 모형에 대한 심도 있는 논의는 지면관계상 추후의 과제로 남겨두겠다.

제외하면, 사회과학 분과학의 학부 및 대학원 필수 커리큘럼에서 수리 모형에 대한 체계적인 교육이 포함된 경우는 매우 드물다. 이와 같은 상황에도 불구하고, 한국의 사회과학 분야에서 통계물리학적 전통에서 비롯한 개념과 연구 결과에 대한 관심은 매우 높다고 할 수 있다. 예를 들어, Barabasi & Albert(1999)의 선호적 연결 모형의 설명 대상이 되는 척도없는 연결선 수 분포의 경우를 보자. 척도없는 분포는 김용학 외(2008)가 한국 언론정보학 분야의 인용 네트워크를 분석한 경우와 같이 경험적 네트워크의 특성으로 발견되거나, 사회 현상에서 척도없는 분포적 특성의 발견이 암묵적으로 척도있는 종모양 분포에 기반해 왔던 사회과학 이론의 전회(transition) 계기가 될 수 있다는 장덕진(2020)의 주장과 같이, 다양한 맥락에서 활용되어 왔다. 하지만 많은 네트워크 생성 모형의 원류가 되는 물리학 이론의 맥락에서, 척도없는 분포를 발현(emerge)시키며, 그 생성 메커니즘으로서 강력한 설명력을 갖는 선호적 연결 모형의 설명구조가 상세히 분석된 적은 없었다.

Barabasi & Albert(1999)를 중심으로 하는 본 논문의 논의는 자연과학적 전통에서 연유한 여러 접근들, 예를 들어서 수리 생태학적 접근을 조직 및 경제 사회학 자료에 적극적으로 이용한 네트워크 동역학 모형 연구인 Saavedra, et al.(2011), 네트워크 구조에 대한 일반 통계 모형으로 연결선 중심의 생성 모형을 제시한 Crane & Dempsey(2018) 등, 사회과학자들에게 유용할 수 있는 경험 분석 방법론을 이해하고 적절히 활용하는 데에도 도움을 줄 것으로 기대한다. 이를 바탕으로 경험 사회과학 연구가 이론 모형 연구의 설명력에 기여할 수 있을 방안에 대해서도 간략히 논한다.

II. 과학적 설명의 형식

1. 법칙 연역적 설명과 포괄 법칙

일상생활을 포함한 다양한 영역에서의 설명은 대부분 현상에 대한 원인을 제시하는 구조를 띤다. 부엌에 있는 철수가 아파하고 있을 때, “뜨거운 냄비에 손을 접촉해서”와 같은 설명은 대다수의 사람들에게 적절한 설명으로 여겨질 것이다. 여기에서의 설명은 상식에 의존하는 형태를 띤다. 즉, “임의의 사람은 뜨거운 냄비에 손

을 접촉하면 아프다”라는 보편적인 관계에 대한 명제가 이러한 설명을 이해 가능하게 만든 것이다. 상식을 포함한 보편적인 관계는 뒤이어 소개할 보편 혹은 포괄 법칙에 대응된다.

과학에 있어서 표준적인 설명의 구조인 법칙 연역적(deductive-nomological) 설명도 이러한 형식을 따른다. 설명을 통한 정당화의 본질적인 형식이라고 할 수 있는 법칙 연역적 구조는 Hempel & Oppenheim(1948)에 의해 정식화되었고, 이후의 논의들 역시 이에 기초해 있다.

Hempel & Oppenheim(1948)은 먼저 피설명항(explanandum), 설명하고자 하는 현상과, 설명항(explanans), 설명을 위한 진술로 설명의 요소를 나누었다. 이를 도시하면 다음과 같은 삼단논법으로, 전형적인 연역 논증의 구조를 띠고 있다.

$C_1, C_2, \dots, C_k$	Statement of antecedent conditions	}	Explanans
$L_1, L_2, \dots, L_k$	General laws		
E	Description of the empirical phenomenon to be explained	}	Explanandum

<그림 1> 법칙 연역적 설명의 구조.

Hempel & Oppenheim(1948)의 첫 번째 그림을 다시 그림.

즉, 사전 조건(antecedent conditions: C)과 보편 법칙(general laws: L)의 연역적 결합을 통해서 피설명항의 경험적 특성(empirical phenomenon: E)이 도출된다. 물체가 낙하하는 이유를 공중에 있었다는 사실(사전 조건)과 만유인력의 법칙(보편 법칙) 때문이라고 하는 것이 전형적인 예이다. 이렇게 피설명항이 설명항에 의해서 귀결될 때, 설명항에 포함된 보편 법칙은 포괄 법칙(covering laws)의 역할을 하며, 단칭 진술인 피설명항을 포섭한다고 말한다.³⁾

법칙 연역적 설명 구조하에서 타당한 설명은 피설명항이 설명항의 연역적 귀결 즉, 항등적(tautological) 진술이어야만 한다. 여기에서 피설명항과 설명항의 관계를

3) 과학 연구에서 설명과 검증(test)은 동전의 양면과 같은 관계를 갖는데, 법칙 연역적 설명을 설명이 아닌 가설 검증의 방향에서 볼 수도 있다. 이 때문에 법칙 연역적 설명은 때때로 가설 연역적(hypothetico-deductive) 설명이라고도 불리는데, 선험적으로 가정되거나 귀납적으로 형성된 법칙은 필연적으로 가설의 지위를 갖기 때문이다. 따라서 모든 과학적 논증에서 법칙은 피설명항에 대한 포괄이 기대되는 보편 명제로서 일시적으로 가정되는 것이라고 할 수 있다.

결과와 원인으로도 지칭할 수 있다.

법칙 연역적 설명은 사회과학 분야에서도 일반적이다. 비록 사회과학에서 법칙의 지위는 자연과학과는 현저히 다르다고 볼 수 있지만, 개별적 설명 상황에서 포괄 법칙에 해당하는 사회과학 이론 혹은 상식을 가정할 수밖에 없다. 예를 들어, 왜 사람들이 일 년 뒤에 15만 원을 받는 것에 비해서 바로 10만원을 받는 것을 선호하는지 설명할 때 행동 경제학의 시간차 할인(intertemporal discounting) 이론이 사용되는데, 행위자 일반의 시간차 할인 경향에 대한 귀납적 믿음은 피설명항을 포괄하는 보편 법칙의 역할을 한다.

2. 인과적 메커니즘에 기반한 설명

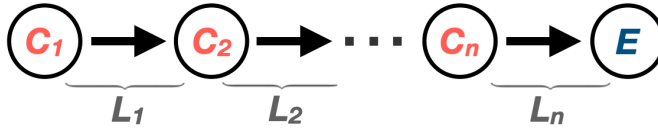
인과적 메커니즘에 기반한 설명은 법칙 연역적 설명의 자연스러운 확장이다. 인간에 대한 이해가 부족한 외계인에게 부엌에 있는 철수가 아파하고 있는 이유를 묻는다면, “뜨거운 냄비에 손을 접촉해서”라는 설명은 충분하지 않다. 이러한 상식이 귀납적으로 형성되지 않은 외계인에게도 이해 가능한 설명은 고통이 전달되는 신경 생물학적 메커니즘을 제시하는 것이다.

그들에게 타당한 설명은 다음과 같다. 철수의 손 상피 세포 아래의 감각 수용체로서 단백질로 이루어진 TRPV1 채널은 역치를 초과하는 온도의 물체와 접촉할 때 열리게 되는데, 여기에서 만들어진 전기 신호가 신경세포(neuron)의 축삭(axon)을 따라 이온 채널이 순차적으로 개폐하면서 만들어진 전압 차이에 의해 전달되고, 신경세포들 사이의 근접 구조인 시냅스(synapse)에서 신경전달물질의 화학적 신호로 매개되어, 다음 신경세포의 가지돌기(dendrite)로 전달된다. 이 전기 신호가 고통을 매개하는 C 그룹 신경 섬유의 길(pathway)을 따라 대뇌변연계의 편도체(amygdala)까지 전달된다. 이러한 연쇄적 전달 과정을 통해 비로소 철수는 고통을 느끼게 된 것이다.

이와 같이 원인과 결과의 연쇄로 이루어진 인과적 고리를 이용한 설명은 개별 고리와 인접한 현상을 포괄하는 부분 집합 각각이 법칙 연역적 설명 구조를 갖는다고 할 수 있다. 동시대 신경과학이 제시하는 이들 각각의 포괄 법칙, 즉 이온 채널이 열리는 전기적 원리 등은 물리화학적 법칙이다.

인과적 메커니즘을 이용한 설명 방식을 인과적 그래프(causal graph)를 통해 나타내면 <그림 2>와 같다. 전체 구조를 매개하는 진술, 사전 조건 C_i 와 포괄 법칙 L_i

의 귀결인 피설명항 E_i 는 다음 연결쌍(dyad)인 $C_{i+1} \rightarrow C_{i+2}$ 혹은 $C_{i+1} \rightarrow E_{i+1}$ 의 사전 조건인 C_{i+1} 가 된다. 그리고 이 단계적 고리는 최종적으로 설명되어야 하는 E 까지 나아간다. 개별적인 연결쌍은 그 각각이 법칙 연역적 설명구조를 띠고 있다.



<그림 2> 인과적 그래프를 통해 나타낸 인과적 메커니즘을 이용한 설명의 일반 구조.

모든 현상은 시간을 축으로 한 순차적 인과 과정으로서 벌어진다. 사회 현상 역시 여러 단계의 인과적 연쇄로서 발생한다. 이미 사회과학 연구에서 타당한 설명의 핵심적 요건으로 인과적 메커니즘에 기반한 설명을 강조한 학자들은 많이 있었다. 사회과학 철학의 대표적인 저작이라고 할 수 있는 Elster(1989), Little(1991), Tilly(2001), Hedstrom(2005), 모두 이러한 견해를 강하게 피력하고 있으며, 이들의 학문적 배경은 철학을 비롯해서 사회학, 경제학, 정치학 등, 사회과학 전 분야를 포괄한다. 메커니즘 분석은 인과적 추론에 대한 통계적 도구가 정교화되면서 경험적 분석 방법론의 영역에서도 활발하게 연구되고 있다(VanderWeele 2015).

뒤이은 논의에서는 메커니즘에 기반한 설명 방식이 본질적으로 물리학을 비롯한 자연과학의 설명 형식과 맞닿아 있다는 점을 이야기하겠다. 이를 바탕으로, 대표적인 두 모형을 중심으로 네트워크 생성 모형의 특성에 대해 설명하겠다.

III. 네트워크 생성 모형 연구의 배경

1. 물리학 이론의 체계

유클리드의 공리 연역적 체계를 경험 분과학의 영역에서 완전하게 구현했다고 여

겨지는 최초의 저작은 뉴턴의 자연철학의 수학적 원리인 《프린키피아》(*The Principia*)이다. 피설명항이 선험적인 진술인 수학, 논리학, 형이상학, 인식론과는 달리, 《프린키피아》는 경험 세계의 투사(projection)로서 연역적 논리 체계를 구성했으며, 피설명항에 해당하는 진술은 경험 세계에서 관찰되는 사물의 운동 현상에 대한 설명력을 제공한다는 점에서 이전의 공리 연역적 체계를 차용한 저작들과는 근본적인 차이가 있다. 《프린키피아》는 이후 물리학 이론의 표준적인 설명 형식 혹은 방법론을 정초했다는 점에서도 혁명적인 저작이라고 할 수 있다. 이 책에서 공준인 힘의 3 법칙을 포함한 공리(공준, 정의, 상식)는 기술된 모든 역학 현상을 설명하는 논리적 출발점으로 이용된다.

설명 구조의 틀에서 보자면, 《프린키피아》 이후, 물리학에서의 설명은 인과적 메커니즘에 기반한 법칙 연역적 구조를 갖게 되었다고 할 수 있다. 관찰된 현상에 대응하는 피설명항을 정당화하는 설명항은 이미 공리적 위상을 갖는 물리학의 근본 법칙에 함축되어 있으며, 이는 설명에 이용된 포괄 법칙의 보편성을 보장한다. 이러한 구조는 통계 물리학의 이론적 접근을 적극적으로 이용한 네트워크 생성 모형 연구들에서도 마찬가지로 발견된다. 공리 연역적 추론은 관찰된 물리 현상을 설명하는 이론 체계의 수학적 정합성을 요구하며, 그 연역 단계 각각이 절차적 타당성을 갖춘다면 모든 이론적 귀결은 공리적 지위를 갖는 진술과 동치이다. 또한, 시간을 따라 진행되는 물리 현상은 그 부분 각각이 물리 법칙으로 설명되는 원인과 결과의 사슬을 구성한다고 할 수 있다. 관찰 가능한 대부분의 물리현상은 어떤 물리적 절차가 본질적으로 단 한 번만 적용되어서 일어날 수는 없기 때문이다. 공리 연역적 이론 구조와 법칙 연역적 메커니즘에 기반한 설명은 물리학을 자연과학 분과학 일련의 포괄 법칙 역할을 담당하는 궁극의 학문으로 만들었다.

2. 네트워크 생성 모형의 기원

네트워크 혹은 그래프를 수학적으로 기술하는 형식론을 처음으로 제시한 연구는 한붓그리기에 대한 오일러의 1735년 논문으로 여겨진다. 이는 이산 수학(discrete mathematics)의 중추적인 분야인 그래프 이론(graph theory)의 탄생으로 이어졌다. 이후 그래프 이론은 경험적으로 관찰되는 개체 간의 관계를 다루는 네트워크 과학에서도 기본적인 표상 틀로서 기능하게 된다(Butts 2009).

현대적인 네트워크 과학의 시발점을 알리는 논문으로는 일반적으로 에어디쉬와

레니의 무작위 네트워크 형성 모형에 관한 논문이 회자된다(Erdos & Reyni 1960). 이 논문과는 독립적으로 동일한 모형이 Gilbert(1959)에 의해 발표된 바 있기 때문에, 이들의 모형을 Erdos-Reyni-Gilbert(ERG) 모형이라고 부른다. ERG 모형은 두 개의 모수(parameter)로 이루어져 있는데, 노드의 개수 N 과 연결확률 p 를 이용해서 $G(N, p)$ 로 나타낸다.

ERG 모형은 N 개의 노드가 존재하는 상태에서 노드 쌍 사이의 연결이 무작위적으로 생성되는 과정을 확률적으로 기술한다. 두 노드를 무작위로 추출해서 연결하는 과정을 반복할 때 생성되는 네트워크 구조의 분포적 특성을 기술하는 것이 이 모형의 목표이다.

주목할 만한 ERG 무작위 네트워크의 분포적 특성은 각 노드별 이웃 수의 분포, 즉 연결선 수(degree) k 의 분포이다. $G(N, p)$ 에서 모든 노드들은 동일하게 p 의 확률로 선택되기 때문에, 임의의 노드가 k 의 이웃을 가질 확률은 이항 분포로 다음과 같이 간단하게 나타낼 수 있다.

$$p(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

대부분의 연구에서 분석 대상이 되는 네트워크는 공통적으로 한 노드의 평균적인 이웃의 수 $\langle k \rangle$ 가 네트워크에 존재하는 전체 노드의 수에 비해서 굉장히 적다는 성김(sparsity) 특성(i.e. $\langle k \rangle \ll N$)을 갖고 있다. 성김 조건하에서 연결선 수 분포는

$\binom{N-1}{k} \approx \frac{(N-1)^k}{k!}$ 라는 근사적 관계를 이용하여 다음과 같이 푸아송(Poisson) 분포로 근사된다.

$$p(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \approx e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

N 과 p 라는 두 개의 모수가 분포의 형태를 결정하는 이항 분포와 달리, 푸아송 분포는 분포 평균에 해당하는 하나의 모수 $\langle k \rangle$ 만을 갖는다. 따라서 이항 분포에 비해서 해석적 연산이 용이하다는 이점을 갖는다. 이 단일 모수 표상(single parameter representation)은 뒤이어 소개할 척도(scale)라는 개념과도 밀접한 연관을 갖는다.

ERG 모형은 통계적 네트워크 모형 연구의 시발점이 되었으며, 추론적(inferential) 네트워크 구조 연구의 기본적인 생성 모형인 지수적 무작위 네트워크 모형(Exponential Random Graph Model: ERGM)의 가장 단순한 형태이기도 하다. 즉, 모든 노드가 모든 상대 노드에 대해서 균질한 생성 확률을 갖는, 회귀분석의 맥락에서는 절편(intercept)만이 존재하는 모형이라고도 할 수 있다.

오랫동안 순수 수학에서 다루어지던 네트워크는 20세기에 들어와서야 인간의 행위나 사회의 특성을 설명하는 형식적 도구로써 적극적으로 활용되기 시작했다. 인류학, 사회학, 사회 심리학 등을 중심으로 상호의존성과 사회체제의 구성적 특성을 연구하는 사회 네트워크 분석이 시작되면서 비로소 네트워크는 경험적으로 관찰 가능한 관계의 집합으로서 접근되었다(Scott 2012). 이에 따라 자연스럽게 네트워크 생성 모형의 설명력은 네트워크의 분포적 특성을 설명하는 이론으로서 사회 네트워크, 기술 네트워크, 생태 및 생물학적 네트워크 등, 실세계 네트워크 전반을 대상으로 시험되기 시작했다.

ERG 모형의 경우, 애초 모형의 주된 목표가 실세계 네트워크의 구조적 특성을 설명하기 위한 것은 아니었고, 일종의 확률 과정으로서 반복 추출을 통해 네트워크의 연결성이 어떻게 변하는지 기술하는 것이기 때문에, 뒤이은 연구에서는 ERG 모형으로 생성한 네트워크와 경험 자료 간의 괴리가 흔히 발견된다.

다음 장에서는 네트워크 과학 역사상 두번째로 많이 인용된 Barabasi-Albert(BA) 모형에서 제안된 선형적 선호 연결(linear preferential attachment) 모형이 ERG 모형이 구현하지 못하는 실세계 네트워크의 어떤 특성을 재현하는지 설명하겠다. 그리고 이 각각의 모형이 법칙 연역적 설명과 검정의 틀 안에서 어떠한 과정을 거쳐서 기각되고 수용되었는지 논한다.

IV. Barabasi-Albert 모형

실세계 네트워크와 이질적인 ERG 모형의 특성은 명확한데, 바로 모든 노드가 균질한, 비선호적 연결을 한다는 점이다. 인스타그램(Instagram)과 같은 온라인 사회 네트워크 플랫폼이나 국가 간 품목별 무역 네트워크와 같은 실세계 자료에서는 혼

하게 평균보다 현저하게 많은 양의 연결을 가진 행위자가 존재한다. 이른바 스타(star) 노드 혹은 허브(hub)의 존재와 역할은 사회 및 임상 심리학자인 모레노(Moreno 1947)에 의해 정식화되었는데, 현대의 정량적 네트워크 연구에서 주목했던 점은 연결선 수의 통계 분포적 특성이었다. 즉, 스타 노드가 평균적인 노드들에 비해서 얼마나 더 많은 연결선 수를 가지고, 어느 정도의 비율로 존재하는지가 관심사였다.

BA 모형의 출발점은 월드와이드웹(World Wide Web: WWW)의 웹페이지 간 네트워크(Albert, Jeong & Barabasi 1999; Adamic & Huberman 2000), 인터넷 라우터 간 네트워크, 영화 공동 출연 네트워크(Barabasi & Albert 1999) 등 다수의 실세계 네트워크의 연결선 수 분포가 ERG 모형으로 생성된 네트워크의 연결선 수 분포인 푸아송 분포와는 명백히 다른 멱함수(power-law)의 형태를 띤다는 사실이었다.⁴⁾ 실세계 네트워크의 연결선 수 분포에 대한 구체적인 정보를 바탕으로 그 생성 과정을 설명하는 모형 개발이 뒤따른 것이다.

멱함수 분포의 다른 이름은 척도없는(scale-free) 분포인데, 이는 통계 물리학의 상전이 이론과 밀접한 연관을 갖는 특성이며, 중요한 동역학적 의미를 갖는다. 직관적으로 이해하기 쉬운 상전이적 특성은 멱함수 연결선 수 분포를 갖는 네트워크는 전염 역치(epidemic threshold)가 존재하지 않는다는 사실이다(Pastor-Satorras & Vespignani 2001). 즉, 표준적인 전염 모형인 감염가능-전염-감염가능(susceptible-infected-susceptible: SIS) 모형하에서 멱함수 분포를 갖는 네트워크는 전염도에 관계없이 바이러스가 자연 소멸하지 않는데, 매우 높은 연결선 수의 허브가 더 쉽게 감염되고 더 많이 전파하기 때문이다. 또 다른 특성으로는 무작위적인 노드 손실(random failure)에서 강건(robust)하다는 점이 있다. 즉, 극소수의 노드만이 광역적인 연결을 갖고, 대부분의 노드들이 매우 낮은 수의 연결을 갖기 때문에, 무작위적으로 노드의 기능이 마비되는 상황에서는 네트워크의 연결성을 비롯한 전체적인 기능이 큰 영향을 받지 않는다. 이와 반대로, 네트워크의 구조적 정보를 바탕으로 허브 노드만을 공격할 경우, 네트워크의 연결성은 파국적인 상황에 처한다(Albert et al. 2000).

4) 사실 BA 모형은 일명 ‘부익부(rich-get-richer)’ 메커니즘을 이용해 멱함수 분포를 생성하는 일반 모형인 사이먼(Simon 1955)의 누적 이득(cumulative advantage) 모형에서 영감을 받아서 논문 인용 네트워크의 구조를 재현한 프라이스(Price 1976)의 모형과 동일하다. 하지만 프라이스의 연구가 서지학 분야 밖에서는 널리 알려지지 않았기 때문에, 바라바시와 알버트에 의해 다시 독립적으로 발견되어 다양한 실세계 네트워크의 구조를 설명하는 모형으로 일반화되었다고 보는 것이 적절하다.

척도가 없다는 표현 그대로의 의미는 변수의 척도 변환(scale transformation)을 해도 분포의 형태가 바뀌지 않는다는 것이다. $p(k) \propto k^{-\gamma}$ 의 형태인 멱함수의 경우 이러한 조건을 수학적으로 만족한다. 척도 없는 분포는 그 평균적인 성질이 분산적 성질에 미미한 영향을 주며, 특히 계의 크기가 아주 커지면, 평균에 상관없이 분산은 발산한다. 이는 ERG 모형에서 푸아송 분포를 따르는 연결선 수 분포의 평균 모수($\langle k \rangle$)가 그대로 분산 모수와 같아지는 것($\sigma_k^2 = \langle k \rangle$) 즉, 척도가 있기 때문에 대부분의 연결선 수가 $(\langle k \rangle - \sqrt{\langle k \rangle}, \langle k \rangle + \sqrt{\langle k \rangle})$ 의 범위에 분포하는 것과는 상반된 특성이다. 멱함수 분포의 경우, 연결선 수 분포의 척도없는 성질, 즉 분포의 표준편차가 평균값에 비해 현저하게 높은 특성은 엄청나게 많은 연결을 갖는 극소수의 스타 노드 혹은 허브 노드에 의해서 야기된다.

멱함수 분포를 갖는 연결선 수를 설명하기 위해 바라바시와 알버트가 제안한 모형(Barabasi & Albert 1999)을 법칙 연역적 설명(<그림 1>)의 형식에 맞게 기술하면 다음과 같다. 사전 조건으로 m_0 개의 노드가 각각의 노드가 적어도 하나의 연결선을 가진 채로 존재한다. 여기에 보편 법칙의 두 가지 요소가 있는데, 성장(growth), 매 시점마다 새로운 노드가 등장하며, 선형적 선호 연결, 새로 등장한 노드는 기존의 노드들이 갖고 있는 연결선 수와 비례하는 확률로 해당 시점에 m 개의 새로운 연결을 형성한다.

이러한 설명항의 요소들이 어떻게 피설명항에 해당하는 멱급수 연결선 수 분포를 가진 네트워크를 생성하는 것으로 귀결되는지 유도할 수 있다. 불연속적인 시간을 가정하는 이 모형을 연속적으로 근사해서, 시간 t 에 따른 임의의 노드 i 의 연결선 수

변화를 표현하면 $\frac{dk_i}{dt} = m \frac{k_i}{\sum_{j=1}^{N-1} k_j}$ 이다. 이미 시간에 따라 노드와 연결선 수가 증가

하는 정도를 설명항의 일부로 특정했기 때문에, 우변의 분모는 $\sum_{j=1}^{N-1} k_j = 2mt - m$

$\approx 2mt$ 로 계산되고, 이를 대입해서 원래의 식을 $\frac{1}{k_i} dk_i = \frac{1}{2t} dt$ 로 변환할 수 있다,

이 식의 양변을 i 가 등장한 시점 t_i 에서 임의의 시점인 t 까지 정적분하면 $k_i(t) =$

$m \left(\frac{t}{t_i} \right)^{\frac{1}{2}}$ 를 얻는다. 매 시점마다 하나의 노드가 생성되기 때문에, 시점 t 에서 특정

상수 값인 k 보다 작은, $k_i(t) < k$ 를 갖는 노드의 개수에 대한 기대값을 $t \left(\frac{m}{k} \right)^2$ 으로

계산할 수 있다. 누적 분포에 대해서 계산된 이 값을 t 가 클 때 전체 노드의 수가

근사적으로 t 와 같다는 사실을 적용해서 $P(k) = 1 - \left(\frac{m}{k}\right)^2$ 로 만들고, 미분을 통해서 k 에 대한 확률 질량 함수로 바꾸면, 연결선 수 멱지수(degree exponent)가 3인 분포에 해당하는 $p(k) = \frac{\partial P(k)}{\partial k} = 2m^2 k^{-3}$ 을 얻게 된다. 즉, 연결선 수가 k 인 노드의 발견 빈도는 k 의 값의 -3 제곱에 비례하며, 이는 k 가 증가함에 따라 그 빈도가 급격하게 줄어든다는 사실을 뜻한다.

이에 따라 연결선 수 분포의 2차 적률($\langle k^2 \rangle = \int_{k_{\min}}^{k_{\max}} k^2 p(k) dk$)은 발산하고, $\sigma_k^2 = \sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2$ 인 분산 역시 발산되는 것을 알 수 있다. 즉, 푸아송 분포와 달리, 멱함수 연결선 수 분포의 평균값 $\langle k \rangle$ 은 분산과 무관하다. 많은 수의 현실 세계 네트워크의 경우, 실제로는 연결선 수 분포의 추정된 멱지수가 2와 3 사이에 존재하는데, 이 경우에는 노드 수 N 이 발산하면, 즉 네트워크의 규모가 매우 커지면, $k_{\max}$ 도 발산하면서, 2차 적률과 분산은 발산하게 된다.

지금까지 BA 모형이 제안된 배경과 그 핵심적 귀결을 살펴보았다. BA 모형은 그 이전의 대표적 네트워크 생성 모형인 ERG 모형이 실세계 네트워크에서 발견되는 구조적 특성인 연결선 수 분포를 설명하지 못한다는 문제를 해결하기 위해서 제안되었으며, 그 핵심적 귀결인 척도없는 연결선 수 분포는 닫힌 형태(closed form)의 계산을 통해, 설명항의 사전 조건과 일반 법칙에서 연역적으로 도출된다. 이제 과학 철학적 맥락에서 BA 모형으로 대표되는 네트워크 생성 모형의 설명적 특성을 알아보자.

V. 네트워크 생성 모형의 설명 구조

1. 피설명항의 특성: 물리량의 특정

앞서 이야기한 바와 같이, 설명 대상인 경험적 현상을 이해하기 위해서 설명 주체는 법칙 연역적 설명을 시도한다. 법칙 연역적 설명의 틀 안에서 타당한 설명은 피설명항인 관찰된 현상에 대한 기술이 설명항의 논리적 함축이다. 즉, 실질 함축에 대한 논리 기호 $\Rightarrow$ 를 이용하면 성공한 설명에 대해서는 설명항과 피설명항 사이에

$\{C, L\} \Rightarrow E$ 의 관계가 성립한다.

이를 바탕으로 ERG 모형의 사례를 보자. 설명항의 사전 조건(C)인 연결선이 존재하지 않는 N 개의 노드와 법칙(L)의 역할을 하는 연결선 형성 규칙이 특정한 구조적 특성을 갖고 있는 네트워크를 만들어 낸다. 만약 피설명항 E 가 ERG 모형의 $\{C, L\}$ 이 함축하는 네트워크의 특성이라면 새로운 이론을 요구할 만한 불일치를 낳지 않았을 것이다.

ERG 모형에서 실세계 네트워크 구조가 피설명항일 때, 이론의 설명력이 도전받는 과정에서 흥미로운 점은 피설명항의 특성이다. 여기에서 피설명항은 일반적으로 네트워크의 구조적 특성을 모두 다 코딩(encode)하고 있다고 말할 수 있는 인접 행렬(adjacency matrix) 혹은 연결 리스트(edge list) 그 자체가 아니다. 그 대신, 비교 혹은 검증되는 경험 네트워크의 여러 구조적 특성 중 하나라고 할 수 있는 연결선 수 분포가 피설명항의 역할을 한다. 만약 이때 A 의 다른 구조적 특성, 예를 들어서 평균적으로 얼마나 많은 이웃을 거쳐야 모든 노드 쌍이 연결되는지를 나타내는 지표인 평균 최단 연결 거리 $\bar{d}(A)$ 와 같은 양이 이용되었다면, 매우 낮은 최단 연결 거리를 갖는 실세계 네트워크의 일반적인 특성과 ERG 모형이 만들어낸 네트워크의 특성이 크게 다르지 않기 때문에, 피설명항 $\bar{d}(A)$ 에 대한 ERG 모형의 함축은 도전받지 않았을 것이다.

이렇게 특정한 측정치, 혹은 분포적 특성을 대상으로 이론을 검증하는 것은 물리학이 물리적 세계를 기술하는 방식과 크게 다르지 않다. 이른바 물리량(physical quantity)이라고 하는 기술 단위는, 그 등장이 자연과학의 특정 분야를 열기도 했다. 뉴턴 역학이 질량, 가속도, 힘 등을 개별 입자에 관한 미시적(microscopic) 물리량으로 정의하면서 출발한 것이 대표적이다. 유용한 물리량의 발굴은 매우 중요한데, 예를 들어, 역학에 있어서 속도와 질량의 곱인 운동량(momentum)은 속도보다 훨씬 더 유용한 물리량이다. 속도를 통한 역학적 현상의 기술이 매우 제한적인 반면, 운동량은 운동하는 입자 집합의 질량 중심을 알 수 있게 하며, 보존 법칙이 성립하는 본질적인 물리량이기 때문이다.

열역학에서 기술의 대상이 계(system)의 온도, 압력과 같은 거시적(macroscopic) 물리량으로 정의된 것 또한 그렇다. 이는 입자의 수가 많을 때 개별적인 입자의 역학적 기술이 불가능함을 간파한 물리학자들의 고민 속에서 등장했다. 거시적 물리량의 제안은 미시적 물리량을 대체하는 개념으로서 입자 다수의 운동을 통계적으로 기술하는 통계역학을 탄생시켰으며, 뒤이어 양자역학과 정보이론과 같은 분야의 탄

생에도 직접적인 영향을 준다.

$\{C, L\} \Rightarrow E$ 의 구조에서 검정 대상이 되는 E 에 해당하는 연결선 수 분포는 네트워크의 중시적(mesoscopic) 혹은 거시적 특성에 해당한다. 이는 미시적 특성이라고 할 수 있는 개별적인 노드 특성과는 대조되며, 직관적 중요성을 갖는 네트워크의 다른 특성 역시 함축한다. 즉, 앞서 이야기한 바와 같이, 연결선 수 분포는 네트워크 상에서의 바이러스 동역학, 견고성과 같은 특성을 결정하는 데에 본질적인 역할을 한다는 점에서 매우 높은 설명력을 갖는 유용한 물리량으로 여겨졌다. BA 모형의 등장에 맞추어 네트워크 과학자들 사이에서 연결선 수 분포의 중요성은 매우 빠르게 공유되었는데, 바라비사와 알버트의 연구에 의해 연결선 수 분포라는 물리량의 엄청난 잠재력이 발굴된 것이라고 할 수 있다.

연결선 수는 사회 네트워크 연구에서도 중시되는 측정량으로, 개별 노드의 권력 혹은 영향력을 나타내는 중심성(centrality) 지표 중 가장 단순한 형태이다. 짐멜(Simmel 2011)과 모레노(Moreno 1947)의 고전적 저작에서 이미 영향력, 권력과 같은 개념으로 표상된 바 있으며, 그 분포 역시 이러한 특성 분포의 불평등 정도를 나타내는 지표로서 분석 대상인 사회 구조의 중요한 일면을 보여준다고 할 수 있다.

2. 설명항과 포섭 방식의 특성: 재귀적 인과 메커니즘으로서의 법칙과 그 적용

앞서 소개한 피설명항의 성격에 더해서, BA 모형의 설명 방식에는 두 가지 특징이 관찰된다. 첫째로 설명항의 특성이다. 여기에서 사전 조건과 보편 법칙의 결합은 일종의 알고리즘(algorithm)을 구성하게 되는데, 확률적으로 기술된 절차를 재귀적(recursive)으로 반복한다. 이를 인과적 메커니즘으로 이해할 수 있는데, <그림 2>의 관계에서 각각의 연결쌍을 포괄하는 법칙 L_i 는 매시간 새로운 노드들의 등장과 선행적 선호 연결을 나타내는 절차가 특정 시점에 적용된 것에 대응된다. 즉 $\forall i, j \ L_i = L_j$ 이다. 초기 조건 C_1 에서 매 시점 동일한 법칙의 적용을 통해 네트워크가 성장하고, 성장한 네트워크는 다음 시점에서 사전 조건으로 기능한다.

이러한 구조는 자연과학 일반의 설명 방식이기도 하다. 당구대 위에서 당구공이 가속도 운동을 할 때 지면의 마찰력이 작용하는 현상, 감각 신호가 이온 채널의 개폐를 따라서 전달되는 현상에 대한 설명 모두 전자기 법칙을 순차적으로 반복 적용한다. 이는 연역적 설명에 적용된 법칙이 동일하며, 매 단계의 법칙이 성립하는 데에 추가적인 가정이 요구되지 않는다는 점에서 물리학 이론 구조의 단순성과도 합

치한다. 마찬가지로, BA 모형을 포함한 네트워크 생성 모형 일반은 뉴턴 역학이 힘의 3 법칙과 같은 최소 원리에서 출발한 것과 같이, 재귀적으로 적용될 수 있는 매우 간략하고 명확한 법칙의 지배를 받는다.

둘째로, 설명항을 통한 피설명항 포섭 방식의 특성을 들 수 있다. BA 모형을 비롯한 수리적 네트워크 형성 모형이 논리적 함축을 드러내는 방식은 앞서 BA 모형이 함축하는 네트워크의 연결선 수 분포에 대한 전체 유도 과정에서 확인할 수 있듯이, 객관적인 절차에 의해 연역적으로 성립한다. 즉, 수학적 유도 능력이 있는 모든 사람은 각자의 주관성이 개입될 여지없이 완전하게 투명한 절차로 $\{C, L\} \Rightarrow E$ 를 완결할 수 있다.

3. 자연과학 전통의 설명과 사회과학 경험 연구의 설명: 과소 결정과 주변화

자연과학적 전통에서 고안된 네트워크 형성 모형의 피설명항, 설명항, 그리고 이들의 연역적 연관 방식은 사회과학 경험 연구 일반의 특성과는 대조를 이룬다.

재귀적 법칙 연역적 설명 덕에 그 메커니즘이 완전하게 투명하다고 할 수 있는 설명 방식은 한편으로 명백한 입증 부담을 갖고 있다. 이는 과학철학에서 과소 결정(underdetermination)이라고 부르는 문제(Stanford 2009), 즉 피설명항을 설명하는 설명항이 유일하지 않을 수 있다는 문제와 연관되어 있다. 타당한 경험적 증거가 뒷받침되지 않는 메커니즘은 대체 메커니즘(alternative mechanism)의 위협에 언제나 노출되어 있다. 물리학의 경우와 같이, 메커니즘의 모든 절차가 투명하게 주장되는 설명은 그만큼 높은 수준의 입증 부담을 갖게 되는 것이다. 다른 한편으로, 충분한 메커니즘을 제시하지 않는 연구는 선행 조건과 피설명항의 연관 관계가 발생하는 원리에 대한 이해를 어렵게 한다.

사회과학 경험 연구가 자연과학과 같은 수준에서 메커니즘에 대한 세밀한 가설을 제시하고, 이에 대한 경험적 검정을 수행하는 경우는 매우 드물다. 최근 들어서 제안된 인과적 매개 분석(causal mediation analysis)과 같은 통계적 메커니즘 분석 기법이 상당히 강한 가정하에서 여러 가지 조건을 만족하는 극도로 제한된 형태의 자료에만 적용 가능한 것에서 알 수 있듯이(VanderWeele 2015), 대다수의 사회과학 경험 연구는 주변화한 효과(marginal effect)를 식별하는(identify) 것에 머문다. 또한 인과적 매개 분석 등을 통해서 매개 변수 몇 개를 특정했다고 하더라도, 네트워크 생성 모형의 경우와 같이 재귀적으로 적용 가능한 법칙 연역적 설명을 완전히

추구하지 않는다. 즉, 인과적 메커니즘 중간의 어느 단계는 반드시 주변화되어 축약된 형태로 주장되거나 검정될 수밖에 없다.

주변화된 효과란 가능한 모든 인과적 매개 과정에 의한 효과가 더해진 효과이다. 주변화가 특정 변수쌍 간의 중간(intermediate) 혹은 매개(mediation) 변수 모두에 대해서 이루어진다면, 추정되는 양은 이 변수쌍에 대한 총 효과(total effect) 혹은 총화 효과(aggregate effect)가 된다. 예를 들어, “국제 관계에서 민주주의 국가들은 전쟁을 할 확률이 현저히 낮다”는 경험적 사실은 국제기구를 통한 중재, 자국 유권자들의 선거에서의 압력 등, 여러 경로로 매개될 수 있다고 추측되는데(Ray 1998), 이러한 모든 매개 경로의 효과를 더한 것이 주변화한 효과이다. 각각의 개별 메커니즘이 어느 정도의 역할을 담당하는지 추정하는 것은 불가능에 가까운 반면, 한 국가쌍이 모두 민주주의 국가인 조건이 그렇지 않은 경우에 비해 최종적으로 전쟁 확률을 얼마나 낮추는지 아는 것은 상대적으로 현저하게 간단하다.

많은 사회과학 경험 연구들이 이렇게 주변화된 효과만을 측정하는 데에 그치는 것에 대해, Elster(1989)의 두번째 장에 기술된 바와 같이 인과적 메커니즘의 블랙박스(black box)를 열지 않고 방치하는 것과 같다는 비판이 있다. 하지만 결국 문제가 되는 것은 경험 자료를 통한 식별 가능성(identifiability)으로 이는 중간 단계의 변수가 이전의 원인 변수와 어떤 통계적 관계를 맺고 있는가로 결정된다(VanderWeele 2015). <그림 2>에 나온 것과 같은 인과적 경로(causal path) 전체를 경험적으로 특정하려면, 식별 가능성을 만족하는 조건이 모든 인과적 연결쌍에 대해서 만족되어야 하기 때문에, 일반적인 사회과학 연구에서 이를 달성하는 것은 불가능에 가깝고, 제한된 조건에서 전체 사슬의 일부가 축약된 모형을 추정할 수 있게 된다.

이러한 양상은 사회과학에 있어서 이론 연구와 경험 연구의 간극을 벌리기도 했다. 즉, 경험적 정당화를 추구하는 연구는 이론이 내포하는 추측성 메커니즘에 대한 입증이 어렵기 때문에, 주변화된 매우 간략한 관계만을 추정하는 것이 일반적인 절차가 되었다. 그리고 경험 연구에서 메커니즘에 대한 설명은 결과 분석 이후 논의(discussion) 단계에서 가설로 제시되는 것이 보통이다. 이에 반해서 경험 자료를 이용한 식별을 하지 않는 이론 연구는 통계적으로 식별 불가능한 관계에 대한 주장을 할 수 있다. 이러한 간극을 지적하고, 이론적 주장과 경험적 식별 간의 긴밀한 교류를 주장한 이론 모형의 경험적 함축(Empirical Implications in Theoretical Models)이라고 명명된 흐름도 있으며(Achen 2002; Granato & Scioli 2004), 대규모 자료 분석의 시대가 도래하면서 합리적 선택이론에 기반한 이론 연구와 인과적 추론에

기반한 경험 연구 사이의 간극이 줄어들 것이라는 낙관적 전망을 제시한 주장도 있다(Monroe et al. 2015).

사회과학과 자연과학 연구 행태의 이러한 차이를 근본적으로 다른 것이라고 보기는 힘들다. 두 분야 모두에서 상충한 과소 결정 문제에서 기인하는 상충(tradeoff)과 선택의 문제가 존재하며, 사회과학에서도 이론의 수정은 경험적 증거와 기존 이론의 불일치로 인해 추동되는 경우가 많다. 하지만 이론적 메커니즘과 경험적 검정의 괴리는 자연과학에 비해 현저하게 높다. 예를 들어 정치행동의 기념비적 연구인 Gerber et al.(2008)에서 무작위 현장 실험을 통해서 사회적 압력이 투표율을 높인다는 주변화된 효과는 검증되었지만, 그 심리적 메커니즘은 특정되지 않았다. 이에 반해서 자연과학에서는 법칙 연역적 메커니즘을 이용한 설명의 완결성을 집요하게 추구한다. 예를 들어, 이른바 ‘말이 안 되는 초자연적 현상(unreasonable occult quality)’이라고 불리며 뉴턴을 굉장한 좌절에 몰아넣었던 만유인력의 원격 작용 문제, 즉 물체의 접촉 없이 운동이 일어나는 것에 대한 설명 메커니즘의 부재가, 프린키피아 출판 228년 뒤에 아인슈타인의 일반 상대성 이론이 질량에 의해서 시공간이 휘어진다는 가설을 제시하면서, 비로소 국소적 작용에 의한 메커니즘으로 해소된 사례가 대표적이다.

이러한 차이가 발생하는 원인의 하나로 조작화에 있어서 재량 혹은 객관성의 문제가 있다. BA 모형을 예로 들자면, 기술 요소가 되는 수학적으로 정의된 모수와 네트워크의 구조적 특성 이외에는 주관성이 개입될 여지가 있는 어떠한 요소도 존재하지 않는다. 네트워크 생성 모형 내에서 수학적인 대상으로 상정된 요소, 예를 들어 노드의 경우, 주관적으로 해석될 여지가 있는 노드의 심리 상태를 고려할 필요는 없다.

이에 반해서 사회과학의 대상이 되는 개념들은 종종 매우 주관적인 조작화의 과정을 거친다. 행복한 정도, 민주주의, 인권과 같은 개념들은 그 측정 방법에 따라 상당한 수준의 차이를 가질 수 밖에 없다. 사회과학에서 개념의 조작화는 측정이론(measurement theory)라는 분야가 따로 만들어질 정도로 연구자의 재량이 많이 개입될 수 있는 분야이다. 최근의 논쟁으로는 전세계의 인권 수준이 평균적으로 상승하고 있느냐는 논의가 있다. 타당한 측정을 위해서는 전문가 조사를 통해서 얻어진 조사 결과의 시간적 추세를 문항반응이론(item response theory) 모형에서 어떻게 통제하느냐는 문제가 있었고, 모형 특징에 따라서 상승 추세의 통계적 유의성이 사라질 수 있다는 사실이 밝혀졌다(Cingranelli & Filippov 2018; Fariss 2019).

이는 친교 네트워크와 같은 사회 네트워크를 측정할 때도 마찬가지인데, 사회적 친밀도를 개념화 하는 데에는 연구 설계의 임의성이 개입될 수밖에 없기 때문이다. 실제로 사회관계를 어떻게 정의하는지에 따라서 경험적으로 측정된 사회관계의 영향력은 매우 달라지는데, 예를 들어서, Lee & Lee(2020)는 청소년기 우울증에 영향을 주는 동료 효과(peer effect)가 사회관계의 종류에 따라서 매우 다르게 나타남을 밝혔다. 이는 연결선이 어떻게 정의되었는지 알 필요 없이, 순수하게 수학적인 대상으로만 다루는 네트워크 생성 모형 연구에서는 고려되지 않는 절차이다.

이에 더해서 표본(sample) 변동성의 문제가 있다. 자연과학의 경우, 서울에서 관찰한 수소 원자가 뉴욕에서 관찰한 수소 원자와 다르게 반응하리라는 고려는 하지 않는다. 이는 네트워크 생성 모형의 경우에도 마찬가지이다. BA 모형과 같은 모형에서, 설명 대상이 실세계 네트워크라고 하더라도, 상당히 다른 개성을 가졌을 노드들은 오직 등장한 시점과 같은 매우 적은 수의 모수로만 완전히 특정된다.

이에 반해서 사회과학의 대상이 되는 인간 혹은 집단은 관심 개념들에 어떠한 영향을 줄지 모르는 상당한 수준의 개별성을 지니고 있다. 이러한 개별성은 교란변수(confounder)로서 통제되지 않을 때 효과 추정에 결정적인 걸림돌이 된다. 따라서 어떠한 표본이 표집되느냐에 따라서 발견 내용의 일반화 정도인 외적 타당성(external validity)은 상당한 정도로 변할 수 있다.

사회과학 연구 설계의 최근 논의들은 표본 변동성에서 연유하는 외적 타당성의 문제를 매우 깊게 다루고 있다(Findley et al. 2021). 이때 외적 타당성은 두 가지 차원에서 판별된다. 일반화 가능성(generalizability)은 표본을 통해 얻은 결과를 표본이 속한 모집단(population)의 경향으로 일반화할 수 있는지, 이전 가능성(transportability)은 이 결과를 표본이 속하지 않은 집단에까지 확장할 수 있는지를 지칭한다. 사회과학 연구에서 이 두 가지 기준은 보편적으로 고려되는데, 예를 들어, 사회과학 경험 연구의 표준적인 절차인 가설 검정에서 통계적 유의성을 확인하는 작업은 그 자체로 표본 추정치를 모집단의 특성으로 일반화가 가능한지 확인하는 절차라고 할 수 있다. 다른 한편으로, 사회과학 연구에서 이전가능성의 문제는 실험 연구를 비롯한 여러 분야에서 제기되고 있다. 그중 대표적인 것은 인류학자인 헨리히(Joseph Henrich)의 일련의 연구로, 그의 연구팀은 인간 행동에 관련된 대부분의 연구들이 WEIRD(Western, Educated, Industrialized, Rich, and Democratic)한, 즉, 서구의 교육수준이 높고, 산업화된, 부유한 민주화된 국가 시민을 이용한 표본을 이용해서 얻어졌으며, 주요 발견이 보편적이지 않음을 밝혔다(Henrich et al. 2010). 이렇듯이

분포적 다양성의 정도가 큰 사회적 개념을 연구하는 사회과학자들은 물리적 보편성에 익숙한 자연과학 기반의 연구들과 다르게 다양한 기준의 연구 타당성을 전면적으로 점검해야한다.

마지막으로 사회과학 법칙의 지위가 문제된다. 이는 앞서 말한 조작화와 샘플 변동성과도 연관되는데, 본질적으로 사회과학에서의 포괄 법칙은 공리 연역적 체계와 같이 최소 원리에 기반한 완결된 이론 체계에서 유도되는 것이 불가능하다. 그보다는 연역적 기반이 부재한 거친 귀납적 지식에 가깝다고 할 수 있다. 이 때문에 자연과학의 법칙에 비해 임시방편적이다. 일단 조작화의 임의성에 취약하다. 특정한 사회과학적 개념에 단일한 이해 혹은 정의는 존재하지 않으며, 개인에 따라 특정한 인과 관계가 발현되는 정도는 크게 다를 수 있다. 또한 귀납적 결론을 도출하는 데에 이용된 표본의 영향을 받는다. 일반적인 경험 연구에서 모든 종류의 교란변수를 통제하는 것은 불가능하고, 무작위 실험마저도 유한한 표본 크기로 인해 통제하지 못하는 교란변수를 갖는다고 할 수 있기 때문에, 본질적으로 통제하지 못한 샘플 특성에 대한 조건적 효과(conditional effect)를 식별한다고 볼 수 있다. 이에 따라 경험 관찰을 통해서 자연과학 수준의 보편 법칙을 도출해내기는 어렵다.

4. 네트워크 생성 모형 연구의 발전 경로와 연구 경향

앞서 ERG 모형에 이은 BA 모형의 등장 사례에서와 같이, 네트워크 생성 모형은 설명형-피설명형 불일치 상황과 과소결정에 대한 문제 제기를 지속적으로 겪으면서 발전해 왔다. 이는 분과학 일반의 발전 과정과도 일치한다고 할 수 있다. 특히 BA 모형의 설명력에 대한 도전과 개선은 꾸준히 이루어져 왔다. 이어지는 논의에서는, 지금까지의 서술을 바탕으로, 관련 분야의 연구가 그동안 어떻게 전개되었는지 소개하고, 앞으로 어떤 방향으로 발전할지 전망하는 데에 도움을 주는 것을 목표로 한다.

과소 결정 문제의 맥락에서, 척도없는 연결선 수 분포를 생성하는 대체 메커니즘을 제안한 연구도 있었다. 복제(copying) 모형(Pastor-Satorras et al. 2003)과 최적화(optimization) 모형(Fabrikant et al. 2002) 모두 BA 모형과 상당히 다른 규칙으로 척도없는 연결선 수 분포를 생성해 냈다. 이들은 각각의 생성 조건 역시 척도없는 연결선 수 분포를 함축한다는 것을 증명했다.

또한 ERG 모형이 함축하는 연결선 수 분포가 실제 네트워크의 분포와 차이가 난다는 관찰이 BA 모형의 등장을 촉발한 것과 같이, 연결선 수 분포 이외에 BA 모

형이 함축하는 네트워크의 구조적 특성이 실세계 네트워크와 상이하다는 점이 지적되었다. 대표적인 특성으로 내 친구 두 명이 서로를 알 확률이 무작위적으로 선택된 두 명이 서로를 알 확률보다 현저히 높다는 군집(clustering) 경향에 대한 관찰이 있다. 이는 군집화된 척도없는 분포(clustered scale-free network)와 같은 새로운 모형의 제안으로 이어졌다(Holme & Kim 2002).

여기에 더해서 직접적으로 BA 모형의 법칙 요소에 대한 검정과 개선도 이루어졌는데, 대표적으로 노드의 연결선 수가 등장 시점과 명확한 상관관계가 없는 경우도 많다는 지적이 있었다. 인스타그램 계정을 언제 개설했는지가 팔로워 수를 잘 설명해주지 않는 것처럼, 시간 이외의 다른 요소 역시 연결선 수의 차이를 설명해주어야 했다. 이를 위해서 적응도(fitness)라는 추가적인 노드 수준 모수가 연결선 수와 함께 선형적 선호 연결 확률을 결정하는 모형이 제안되었다(Bianconi & Barabasi 2001).

최근에는 실세계 네트워크 중 실제로 척도없는 연결선 수 분포를 갖는 네트워크가 얼마나 되는지에 대한 논쟁이 있었다. 이는 BA 모형의 경험 자료에 대한 설명력에 의 도전이었다고 할 수 있다. Broido & Clauset(2019)은 천 여 개의 실세계 네트워크의 연결선 수 분포의 함수적 형태를 멱함수와 통계적으로 비교·검정하여 이들 중의 4%만이 실제로 척도없는 분포를 보임을 밝혔다. 하지만 이는 앞서 소개한 바와 같이, 네트워크의 노드 수가 무한에 이를 때만 척도없는 연결선 수 분포 특성이 모든 연결선 수에 대해 관찰되기 때문에, 유한한 크기의 네트워크의 경우에는 매우 큰 연결선 수 영역에서 지수적 분포와 유사한 분포를 보이는 지수적 탈락(exponential cutoff) 현상에서 기인했을 뿐이라는 반발이 나왔다(Holme 2019; Serafino et al. 2021).

많은 수의 네트워크, 특히 사회 네트워크는 Newman & Park(2003)에 의해 분석된 바와 같이, 그 대부분이 척도없는 연결선 수 분포 특성을 갖지 않으며, BA 모형이 생성할 수 없는 보편적인 구조적 특성을 갖는다. 이러한 구조적 특성이 발견되는 이유로 네트워크 생성 과정에서 이질적인 개인 특성이 중요하다는 사실이 부각되어 왔으며, 지수적 그래프 모형으로 일반화되어 다루어지고 있다(Lusher et al. 2013). 여기에서 각각의 노드쌍은 사회적 특성 변수를 포함한 일반화된 변수 관계로 서로를 잇는 연결선 생성 확률이 결정된다. 또한 이러한 특성을 잠재적 변수, 즉 추론 대상으로 삼는 모형들도 제안되었는데, 잠재적 변수를 연속 변수로 특정한 잠재적 공간 모형(latent space model)(Hoff et al. 2002)과 불연속적 변수로 특정한 확

률적 블락모형(stochastic blockmodel) (Abbe 2017; White et al. 1976)이 대표적이다. 이러한 모형들은 네트워크 구조 모사(simulation)을 위한 네트워크 생성 모형뿐만 아니라, 회귀분석 계수, 잠재적 변수 추정을 위한 통계적 추론 방법론의 생성 모형(generative model)으로서 적극적으로 이용되고 있다.

최근의 사회 네트워크 연구에서는 분과학 구분 없이 여러 종류의 방법론과 이론적 관점이 공유되고 있다. 특히 사회과학 분야에서 널리 알려진 미시적 메커니즘을 통해 네트워크가 생성되는 과정을 상정하고, 경험 자료 분석을 통해서 메커니즘에 이용된 모수를 직접 추론하는 형태의 분석 모형도 많이 발표되고 있다. 대표적으로 네트워크 게임 이론의 균형 개념에 기반한 전략적 연결선 형성에 대한 일련의 통계적 구조 모형(structural model)이 있다. 예를 들어, de Paula et al.(2018)은 관찰된 네트워크가 행위자들의 선호적 연결에 따른 전략적 균형(pairwise stable) 상태라고 가정하고, 잠재적 선호 구조를 추론하는 일반 모형을 제안한 바 있다.

네트워크 구조에 대한 조사 방법론 개발도 성행하고 있는데, 그중 하나가 매우 크거나 표집이 어려운 네트워크의 특성 추정 문제이다. 이러한 연구들은 표본 특성을 통해서 모집단의 특성을 높은 정확도로 측정하는 목표를 갖는데, 특정 인구 그룹에 속한 사람을 몇 명 아는지 묻는 문항 설계를 바탕으로 실제 사회 네트워크의 연결선 수 분포를 추론하는 McCormick et al.(2010)의 방법론이 대표적이다. 이와 같은 창의적인 조사 방법론을 활용하지 않고 실세계 사회 네트워크의 구조를 추론하는 것은 매우 어려운 일이다. 이렇게 추론된 연결선 수의 분포적 특성을 통계적 네트워크 생성 모형에 기반한 방법론, 예를 들어 주어진 연결선 수 분포를 갖는 모든 네트워크 구조에 대한 조합론적 분석을 한 통계물리학 논문인 Anand et al.(2014)의 결과를 바탕으로 재구성한다면, 연결선 수 분포와 같이 조사연구를 통해 얻은 네트워크 구조에 대한 부분적 지식을 다른 구조적 특성에 대한 추론적 지식으로까지 확장할 수 있을 것이다.

VI. 나가면서

분과학 간 협업의 중요성과 학제적 연구 경향이 대두되고 있지만, 단일한 분과학

의 커리큘럼을 습득한 대부분의 학자들이 상대 학문의 전통하에서 생산된 연구를 이해하는 것은 쉽지 않은 일이다. 이는 다양한 분과학 출신의 학자들이 유입된 계산사회과학 분야에서와 같이, 분석 대상이 되는 자료가 대동소이하더라도 힘든 일인데, 흔히 “그래서 뭐?(So what?)”와 같은 반응이 심심치 않게 따라다니고, 논문의 입증 혹은 정당화 방식이 납득되지 않는 경우도 많다.

본 논문은 이를 해소하는 것은 타 분과학의 연구가 기초하고 있는 설명의 구조를 이해하는 데에서 시작할 수 있다는 믿음에서 쓰여졌다. 지금까지 대표적인 학제적 분야인 네트워크 과학의 널리 알려진 네트워크 생성 모형을 바탕으로, 그 설명 구조가 어떠한 전통에서 비롯되었는지 분석하고, 사회과학 경험 연구 구조 일반과의 대비를 통해서 서로 간의 공통점과 차이점을 알아보았다. 이 과정에서 피설명항과 설명항의 특성과 연역 방식에 있어서의 차이가 드러났으며, 이는 생성 모형에서 물리학 전통의 재귀적 법칙 연역적 설명 방식과 메커니즘에 대한 강조가 특정 물리량에 대한 경험 검증으로 이어진 결과임을 주장했다.

주지해야 할 것은 이 논문에서 자연과학 전통 혹은 사회과학 전통의 방법론 중 특정 접근이 우월하다는 주장을 하고 있지 않다는 것이다. 이 두 접근은 보완적 역할을 해야 하는데, 예를 들어, BA 모형을 수용한다고 해도 수리적인 접근만으로는 선형적 선호 연결이 일어나는 이유를 알 수 없다. 그 이유를 특정하는 것은 BA 모형의 연역 구조가 포괄하지 못하는 인과적 메커니즘의 선행 층(preceding layer)을 추가한다. 이는 네트워크 생성 모형상의 연결 주체가 되는 행위자들의 효용 함수를 특정해서 현실 네트워크에서의 연결선 형성 과정을 합리화(rationalize)하는 네트워크 경제학 분야의 연구 경향과도 합치한다(de Paula et al. 2018). 즉, 사회과학자들에게 익숙한 개념인 심리적 기제, 사회화, 혹은 경제적 효용이 행위자의 선호적 연결 경향을 설명함으로써 더 완전한 인과적 메커니즘을 규명하는 것이다. 이와 더불어, 본문에 소개된 예와 같이, 인간의 심리적 특성에 대한 이해와 조사 방법론에 대한 실제적 지식은 네트워크 자료 수집 단계에서 표본 편향을 줄이는 데에 기여할 수 있다.

학제적 배경을 가진 연구자가 늘어나고, 분과학 간의 교류가 더욱 빈번해지면서, 사회과학자와 자연과학자가 서로의 방법론을 활용할 기회는 더 많아질 것으로 기대한다. 이에 따라, 상대의 분야에서 이미 발견된 바를 알지 못하고 중복된 연구를 하거나, 더 효율적인 방법론의 존재를 인지하지 못하는 사례는 더욱 줄어들게 될 것

이다. 이를 통해서 사회과학자들은 사회 네트워크를 비롯한 사회 현상의 구조와 메커니즘을 좀 더 심도 있게, 다면적으로 이해할 수 있다. 이러한 총체적인 이해를 바탕으로, 타 분야 전통의 설명 역시 포괄할 수 있는 이론화를 추구할 것으로 기대한다. 이 논문 역시 원활한 이해와 소통을 위한 작은 역할을 할 수 있길 기대한다.

한정된 지면 때문에 자연과학과 사회과학의 여러 차이에 대해서 모두 포괄하지는 못했다. 특히, 확률적 현상에 대한 통계적 설명이 결정론적 현상에 대한 설명과 어떻게 다른지(Salmon 1971), 또한 베이즈주의 입증 이론(Bovens & Hartmann 2004)과 같은 검정 이론의 관점에서 사회과학과 자연과학의 법칙과 검정의 특성에 대해서는 논하지 못했다. 이는 사회과학의 논증 방식에 집중한 다음 연구를 위한 주제로 남겨 두겠다.

참고문헌

- 김용학·김영진·김영석. 2008. “한국 언론학 분야 지식 생산과 확산의 구조” 《한국언론학보》 52(1): 117-140.
- 장덕진. 2020. “사회과학 이론을 다시 써야 하는가?: 역함수 사회의 도전” 《문명과 경계》 (3): 78-108.
- Abbe, E. 2017. “Community Detection and Stochastic Block Models: Recent Developments.” *The Journal of Machine Learning Research* 18(1): 6446-6531.
- Achen, C.H. 2002. “Toward a New Political Methodology: Microfoundations and ART.” *Annual Review of Political Science* 5(1): 423-450.
- Adamic, L.A. and B.A. Huberman. 2000. “Power-law Distribution of the World Wide Web.” *Science* 287(5461): 2115-2115.
- Anand, K., D. Krioukov, and G. Bianconi. 2014. “Entropy Distribution and Condensation in Random Networks with a Given Degree Distribution.” *Physical Review E* 89(6): 062807.
- Albert, R., H. Jeong, and A.L. Barabási. 1999. “Diameter of the World-wide Web.” *Nature* 401(6749): 130-131.
- Albert, R., H. Jeong, and A.L. Barabási. 2000. “Error and Attack Tolerance of Complex Networks.” *Nature* 406(6794): 378-382.

- Barabási, A.L. 2003. *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. Basic Books.
- Barabási, A.L. and R. Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286(5439): 509-512.
- Bianconi, G. and A.L. Barabási. 2001. "Competition and Multiscaling in Evolving Networks." *Europhysics Letters* 54(4): 436.
- Blau, P.M. (Ed.). 1975. *Approaches to the Study of Social Structure*. Free Press.
- Bovens, L. and S. Hartmann. 2004. *Bayesian Epistemology*. Oxford University Press.
- Broido, A.D. and A. Clauset. 2019. "Scale-free Networks Are Rare." *Nature Communications* 10(1): 1017.
- Butts, C.T. 2009. "Revisiting the Foundations of Network Analysis." *Science* 325(5939): 414-416.
- Cingranelli, D. and M. Filippov. 2018. "Are Human Rights Practices Improving?" *American Political Science Review* 112(4): 1083-1089.
- Crane, H. and W. Dempsey. 2018. "Edge Exchangeable Models for Interaction Networks." *Journal of the American Statistical Association* 113(523): 1311-1326.
- de Paula, Á., S. Richards-Shubik, and E. Tamer. 2018. "Identifying Preferences in Networks with Bounded Degree." *Econometrica* 86(1): 263-288.
- Elster, J. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge University Press.
- Fabrikant, A., E. Koutsoupias, and C.H. Papadimitriou. 2002. "Heuristically Optimized Trade-offs: A New Paradigm for Power Laws in the Internet." in *Automata, Languages and Programming: 29th International Colloquium, ICALP 2002* Málaga, Spain, July 8-13, 2002 Proceedings 29 (pp. 110-122). Springer Berlin Heidelberg.
- Fariss, C.J. 2019. "Yes, Human Rights Practices Are Improving over Time." *American Political Science Review* 113(3): 868-881.
- Findley, M.G., K. Kikuta, and M. Denly. 2021. "External validity." *Annual Review of Political Science* 24: 365-393.
- Gerber, A.S., Green, D.P., and C.W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-scale Field Experiment." *American Political Science Review* 102(1): 33-48.
- Granato, J. and F. Scioli. 2004. "Puzzles, Proverbs, and Omega Matrices: The Scientific and Social Significance of Empirical Implications of Theoretical Models (EITM)." *Perspectives on Politics* 2(2): 313-323.
- Hedstrom, P. 2005. *Dissecting the Social: On the Principles of Analytical Sociology*.

- Cambridge University Press.
- Hempel, C.G. and P. Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15(2): 135-175.
- Henrich, J., S.J. Heine, and A. Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33(2-3): 61-83.
- Hoff, P.D., A.E. Raftery, and M.S. Handcock. 2002. "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association* 97(460): 1090-1098.
- Holme, P. and B.J. Kim. 2002. "Growing Scale-free Networks with Tunable Clustering." *Physical Review E* 65(2): 026107.
- Holme, P. 2019. "Rare and Everywhere: Perspectives on Scale-free Networks." *Nature Communications* 10(1): 1016.
- Imbens, G.W. and J.D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 467-475.
- King, G., R.O. Keohane, and S. Verba. 2021. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.
- Lazer, D., A. Pentland, L. Adamic, S. Aral, A.L. Barabási, D. Brewer, ... and M. Van Alstyne. 2009. "Computational Social Science." *Science* 323(5915): 721-723.
- Lee, D. and B. Lee. 2020. "The Role of Multilayered Peer Groups in Adolescent Depression: A Distributional Approach." *American Journal of Sociology* 125(6): 1513-1558.
- Little, D. 1991. *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science*. Westview Press.
- Lusher, D., J. Koskinen, and G. Robins(Eds.). 2013. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press.
- McCormick, T.H., M.J. Salganik, and Zheng, T. 2010. "How Many People Do You Know?: Efficiently Estimating Personal Network Size." *Journal of the American Statistical Association* 105(489): 59-70.
- Moreno, J.L. 1947. "Contributions of Sociometry to Research Methodology in Sociology." *American Sociological Review* 12(3): 287-292.
- Morgan, S.L. and C. Winship. 2015. *Counterfactuals and Causal Inference*. Cambridge University Press.
- Monroe, B.L., J. Pan, M.E. Roberts, M. Sen, and B. Sinclair. 2015. "No! Formal Theory, Causal Inference, and Big Data Are not Contradictory Trends in Political Science."

- PS: Political Science & Politics* 48(1): 71-74.
- Newman, M.E. and Park. 2003. "Why Social Networks Are Different from Other Types of Networks." *Physical Review E* 68(3): 036122.
- Okasha, S. 2002. *Philosophy of Science: A Very Short Introduction* (Vol. 67). Oxford Paperbacks.
- Pastor-Satorras, R. and A. Vespignani. 2001. "Epidemic Spreading Sin Scale-free Networks." *Physical Review Letters* 86(14): 3200.
- Pastor-Satorras, R., E. Smith, and R.V. Solé. 2003. "Evolving Protein Interaction Networks through Gene Duplication." *Journal of Theoretical Biology* 222(2): 199-210.
- Price, D.D.S. 1976. "A General Theory of Bibliometric and Other Cumulative Advantage Processes." *Journal of the American Society for Information Science* 27(5): 292-306.
- Ray, J.L. 1998. "Does Democracy Cause Peace?" *Annual Review of Political Science* 1(1): 27-46.
- Salmon, W.C. 1971. *Statistical Explanation and Statistical Relevance* (Vol. 69). University of Pittsburgh Press.
- Saavedra, S., D.B. Stouffer, B. Uzzi, and J. Bascompte. 2011. "Strong Contributors to Network Persistence Are the Most Vulnerable to Extinction." *Nature* 478.7368: 233-235.
- Scott, J. 2012. *What Is Social Network Analysis?* Bloomsbury Academic.
- Serafino, M., G. Cimini, A. Maritan, A. Rinaldo, S. Suweis, J.R. Banavar, and G. Caldarelli. 2021. "True Scale-free Networks Hidden by Finite Size Effects." *Proceedings of the National Academy of Sciences* 118(2): e2013825118.
- Simmel, G. 2011. *Georg Simmel on Individuality and Social Forms*. University of Chicago Press.
- Stanford, K. 2009. "Underdetermination of Scientific Theory." in *Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), Edward N. Zalta (ed.),
<<https://plato.stanford.edu/archives/win2009/entries/scientific-underdetermination/>>
- Tilly, C. 2001. "Mechanisms in Political Processes." *Annual Review of Political Science* 4(1): 21-41.
- VanderWeele, T. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.
- Watts, D.J. and S.H. Strogatz. 1998. "Collective Dynamics of Small-world Networks." *Nature* 393(6684): 440-442.
- White, H.C., S.A. Boorman, and R.L. Breiger. 1976. "Social Structure from Multiple

Networks. I. Blockmodels of Roles and Positions.” *American Journal of Sociology* 81(4): 730-780.

<접수 2023.01.31; 수정 2023.02.10; 게재확정 2023.02.24>

The Structure of Explanations in Network Formation Models: An Analysis of Barabasi-Albert Preferential Attachment Model

Yunkyu Sohn
(Seoul National University)

Network formation models proposed around the year 2000 played a key role in establishing network theory as a universal analytical framework for various branches of science, including social science. However, most network formation models are based on a theoretical foundation rooted in natural science, particularly physics, which may be unfamiliar to social scientists. By incorporating a standard perspective in philosophy of science, I analyze the explanatory structure of a network formation model, which is widely acknowledged as the origin of contemporary network science, and compare its explanatory structure with the structure of explanations in social science. Through this, I aim to promote shared understandings among scientific disciplines harnessing social data and discuss the potential contributions of social science to the interdisciplinary trends in academic research.

Key words: network formation model, deductive-nomological explanation, causal mechanism, natural science, social science