

척도 없는 연결망은 드문 걸까 흔한 걸까?

2019년 5월 29일

김범준



얼마 전 출판된 한 논문[1]이 척도 없는 연결망에 대한 논쟁에 다시 불을 붙였다. 거의 천개에 육박하는 현실의 여러 연결망에 통계적 검증방법을 적용해서, 이 중 딱 4%만을 엄밀한 의미에서 척도 없는 연결망으로 볼 수 있다는 것을 밝힌 연구다. 많은 연결망 연구자의 생각은 달랐다. 자연과 사회에 존재하는 수많은 현실의 연결망 중에 척도 없는 연결망이 이보다는 훨씬 더 흔하게 발견될 것으로 믿었다. 깜짝 놀랄 결론이 담긴 논문[1]이 연결망 연구 분야에서 현재 널리 회자되고 있는 이유다(심지어 논문[1]은 출판시점에 이미 50번 이상 다른 논문에서 인용되었다[2]). 첫 문장에서 필자가 '다시'라고 적은 이유가 있다. 이 논문의 저자 클로셋 Aaron Clauset의 2009년 논문[3]도 비슷한 결론을 주장했기 때문이다. 이번 연구[1]는 과거의 연구[3]보다 훨씬 더 많은 연결망을 분석에 이용했다는 점에서, 척도 없는 연결망이 현실에서 아주 드물다는 결론에 더 큰 무게가 실렸다. 논문[1]에 대한 비평comment을 담은 홀메 Petter Holme의 논문[2]이 나란히 함께 출판되었다. [1]의 제목은 "척도 없는 연결망은 드물다Scale-free networks are rare"고, 이 논문에 대한 비평논문[2]의 제목은 "드물지만 어디에나 있다Rare and everywhere"이다. 두 논문 모두 제목이 각각의 결론을 함축적으로 보여준다. 두 논문의 내용이 요약되어 있는 쿼타매거진Quanta Magazine의 기사[4]를 추천한다. 최근의 척도 없는 연결망 논쟁에 대한 각 연구진영의 입장이 잘 정리되어 있다.

'척도 없음scale-freeness'은 통계물리학에서 가장 중요한 개념 중 하나다. 연속 상전이가 일어나는 임계점 부근에서 자주 발견되는 특성이다. 거칠게 설명하자면, 상전이가 일어나고 있는 물리계는, 가까이서 보든 멀리서 보든 똑같이 보인다는 얘기가. 이를 수학적으로 기술하는 것이 바로 노벨상을 받은 윌슨의 되틀맞춤재규격화군 이론이다. 상전이점은 거리의 척도를 바꾸는 재규격화군 변환의 부동점fixed point이다. 상전이점에서는 계를 가까이서 보나 멀리서 보나(즉, 재규격화군 변환을 하면), 똑같이 보인다부동점은 말이다. 통계물리학의 바로 이 척도 없는 임계성은 또 다른 중요 개념인 보편성universality의 근거가 되기도 한다. 가까이서 보나 멀리서 보나 같아 보이니, 가까이서 봤을 때만 보이는 세세한 차

이는 임계현상에 중요한 영향을 미칠 리가 없다고 할 수 있다. 실제로도 물리계가 놓인 공간과 동역학적 변수가 각각 몇 차원인지, 물리계가 어떤 대칭성symmetry을 갖는지와 같은 몇 특성이 같기만 하면, 완전히 다른 물리계라도 정확히 같은 임계현상을 보여준다. 예를 들어 3차원에서 자석이 자성을 갖게 되는 상전이는 기체가 액체로 변하는 상전이와 정확히 같은 임계현상을 보여준다. 세상에 존재하는 구체적인 물리계의 상전이는 정말 다양하지만, 이들 상전이가 단지 몇 개의 보편성류universality class로 무리지어 묶여 구별된다. 통계물리학은 이처럼 수많은 개별자를 관통하는 보편적인 특성에 주목하고자 하는 학문특성이 있다. 다른 것들 사이에서 무엇이 같은지를 보려한다.

척도, 잣대, 혹은 축척의 뜻을 갖는 영어단어 scale에 “~이 없는”의 뜻을 가진 free를 붙인 단어 scale-free는 우리말로는 보통 “척도 없는”으로 번역한다. “척도 없음”은 수학적으로 명확히 정의할 수 있는 특성이다. 자, $y=f(x)$ 라는 함수가 있다. 이 함수에 소위 척도변환scale transformation을 해 볼 수 있다. 척도변환은 물리학에서 단위를 바꾸는 변환에 해당한다. 거리 x 를 미터로 재다가 인치로 재면, 이는 x 를 x/a 로 바꾸는 변환이 된다. 마찬가지로의 척도변환을 y 에 대해서도 하면 y 는 y/b 로 바뀌 적을 수 있다. 즉, 원래의 식 $y = f(x)$ 에 척도변환을 하면 $y/b = f(x/a)$ 가 된다. 척도변환에 대해 불변이라는 말은, $bf(x/a) = f(x)$ 를 만족하는 b 가 임의의 a 에 대해 존재한다는 뜻이다. 척도변환에 대해 불변일 때 척도가 없다고 말한다. 척도를 바꿔도 같아 보려면, 함수 자체에 특정한 척도가 처음부터 없어야 하기 때문이다. 따라서, $y = f(x)$ 가 위의 척도변환에 대해 불변이면 “ $y = f(x)$ 는 척도가 없다”고 이야기한다. 이 식에 $f(x) = Ax^c$ 를 대입하고 $b = a^c$ 를 택하면, 척도변환에 대한 위의 불변 조건을 만족함을 쉽게 확인할 수 있다. 즉, 거듭제곱함수^{멱함수}라고도 부른다 꼴을 따르는 함수는 척도변환에 불변이고, 따라서 척도가 없다. 거듭제곱함수만 척도가 없다 보니, “척도 없음”은 “거듭제곱함수^{멱함수}”와 늘 함께 거론된다.

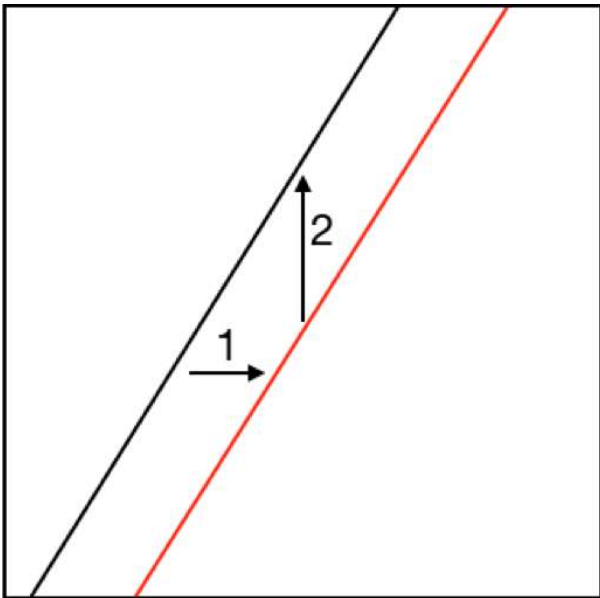


그림1 로그 축척으로 그리면 거듭제곱 함수는 곧은 직선의 꼴이 된다. 가로축 방향으로 평행이동(1번 화살표)하고 나서 다음에 세로축 방향으로 적당한 평행이동 (2번 화살표)을 하면 원래의 직선에 겹쳐지게 할 수 있다. 거듭제곱 함수만 이렇게 할 수 있으니, 거듭제곱함수만 척도가 없다. 직선이 아닌 곡선을 아무거나 그리고는 위의 방법으로 다시 겹쳐지게 만들 수 있는지 살펴보라.

그래프로 함수를 그려 척도 없음을 쉽게 확인할 수도 있다. $y = f(x)$ 를 그래프로 그릴 때, 가로축의 x 와 세로축의 y 를 모두 로그의 축척으로 그려보자. $y = f(x) = Ax^c$ 로 주어지는 거듭제곱함수의 경우, 양변에 로그를 취하면 $\log(y) = \log(A) + c \cdot \log(x)$ 이므로, $\log(y) = Y$, $\log(x) = X$ 로 치환하면 $Y = A' + c \cdot X$ 의 꼴이 된다. 즉, 거듭제곱의 지수가 c 인 거듭제곱함수를 로그의 축척으로 그리면 기울기가 c 인 직선 꼴이 된다. 앞에서 소개한 척도변환에서 x 를 x/a 로 바꾸는 변환은 로그 축척에서는 그래프를 가로축으로 평행 이동하는 것에, 또, y 를 y/b 로 바꾸는 변환은 세로축으로 평행 이동하는 것에 해당한다. 나눗셈 x/a 와 y/b 는 로그 축척에서는 뺄셈 $\log(x) - \log(a)$, $\log(y) - \log(b)$ 에 각각 대응함을 생각하면 쉽게 이해할 수 있다. 이제 이렇게 그린 그림에서 주어진 함수가 척도변환에 불변이라는 말은, 그래프를 좌우로 이동(x 를 x/a 로 변환)한 다음에, 위아래로 움직여(y 를 y/b 로 변환) 원래의 그래프와 다시 겹치게 할 수 있다는 말과 정확히 같다. 독자도 한번 해 보시라. 로그의 축척으로 그린 그래프에서 곧게 뻗은 일직선 그래프(즉 $f(x)$ 가 거듭제곱 꼴일 때)만 가로축 방향과 세로축 방향의 한 번씩의 평행이동에 의해 다시 겹쳐지게 만들 수 있다[그림1]. 물리학자가 어떤 함수가 척도가 있는지 없는지를 판단할 때 가장 먼저 시도해 보는 방법이 있다. 가로축, 세로축 모두를 로그의 축척으로 그려서 곧게 뻗은 직선을 따르는 영역이 있는지를 살핀다.

연결망에서 $y = f(x)$ 에 등장하는 x 와 y 가 무엇인지 알아야 척도 없는 연결망의 의미가 명확해진다. 이 분야 연구자들 사이에서는 x 로는 연결망의 노드가 가진 연결된 이웃의 숫자 k 를, 그리고 y 로는 확률분포 P 를 택하는 것이 이쪽 업계의 합의된 약속이다. 따라서, 척도 없는 연결망이란, 구체적으로는 이웃수 k 의 확률분포가 거듭제곱함수 꼴, 즉 $P(k) \sim 1/k^a$ 를 만족하는 연결망이다. 최근의 논문[1]과 [2]는 바로 이처럼 이웃수 분포가 거듭제곱의 꼴을 명확히 따르는 연결망이 현실에서 얼마나 자주 발견되는지에 대한 논의이다. 척도 없는 연결망의 의미는 수학적으로야 아무런 문제 없이 잘 정의되지만, 현실의 연결망이 정말로 척도가 없는지를 확인하는 것은 정말 어렵다. 여러 이유가 있다. 먼저, 거듭제곱 꼴을 잘 따르는 이웃수 k 의 영역이 수학적으로야 1에서 무한대까지지만, 현실에서는 여러 이유로 그럴 수 없어 $k_{\min} < k < k_{\max}$ 의 영역에서만 거듭제곱 꼴이 관찰된다. 이 부등식에서 k_{\min} , k_{\max} 를 얼마나 해야 할지, 또, 얼마나 정확히 거듭제곱의 직선 꼴을 따라야 척도가 없다는 결론을 내릴 수 있는지를 정하는 것은 무척 어려운 문제다. 통계학에서는 귀무가설을 설정하고 이를 기각할 수 있는지를 엄밀하게 살펴보는 방법이 잘 발달되어 있다. 논문[1]은 이러한 엄밀한 통계학적인 방법을 적용해서 현실의 데이터를 가지고 만든 실제의 연결망 중 이 테스트를 통과해, 강한 의미로 척도 없는 연결망이라 할 수 있는 연결망이 극히 소수라는 것을 보인 연구다.

//

필자의 생각도 홀메의 논문 제목

“Rare and everywhere”와 비슷하다.

보기 어렵다고 해서, 정말로 드물다는 결론을 곧바로 내릴 수는 없다.

//

우리말 속담에 “팔은 안으로 굽는다”고 했다. 통계학적인 방법에 기반한 논문[1]의 “척도 없는 연결망은 아주 드물다”는 결론을 물리학자인 필자가 액면 그대로 받아들이기는 어려웠다. 논문[1]의 통계적 방법에 심각한 문제가 있다는 뜻이 전혀 아니다. “많은 연결망은 척도가 없는 거듭제곱 꼴의 이웃수분포를 보여준다”말이 물리학에서 어떤 의미를 가지는지를 생각하면, 논문[1]의 결론은 비록 맞더라도 큰 의미가 없어 보인다.

퀀타매거진의 글[4]에서 물리학의 배경을 가진 연결망 연구자 바라바시^{Barabási Albert-László}도 비슷한 생각을 피력한다. 즉, 논문[1]은, 크고 작은 돌멩이를 떨어뜨리는 실험을 여러 번 하고 그 데이터를 분석해서, 어떤 돌도 엄밀한 의미에서 자유낙하를 하지 않는다는 결론을 내리는 것과 비슷하다는 입장이다. 복잡한 현상을 설명할 때 그 현상에 가장 중요한 영향을 미치는 요인에 우선 주목하는 물리학의 어림의 방법론을 생각하면 논문[1]의 결과는 사실 별로 놀랍지 않다. 오히려, 4%에 해당하는 실제 연결망이 엄밀한 의미로도 척도 없는 연결망으로 볼 수 있다는 통계적인 결과가 더 놀랍다고 할 수도 있다. 이는 마치, 공기 저항, 부는 바람, 지구의 자전에 의한 코리올리의 효과 등이 있는 데도 불구하고, 100개의 돌을 여기저기서 낙하시켰더니, 물체의 속도가 엄밀하게 $v = gt$ 를 따르는 돌이 4개가 있었다는 얘기처럼 들렸다. “딱 4개뿐”을 “무려 4개나”라고 해석할 수도 있다는 입장이다. 마찬가지로 물리학의 배경을 가진 홀메도 논문[2]에서 비슷한 의견을 제시한다. 엄밀한 의미의 척도 없음은 크기가 무한대인 연결망에서만 관찰할 수 있으므로, 어쩔 수 없는 유한성으로 제한된 현실의 연결망 분석을 통해 진정한 의미의 척도 없음을 직접적으로 확인하기는 어렵다는 주장이다. 물리학에서 어떤 현상을 설명하는 여러 요소가 있을 때, 가장 중요한 요소만을 고려하고, 이에 기반해 차근 차근 한 단계씩 나아가는 방법을 자주 쓴다. 필자는 “척도 없음”의 입장에서 현실의 연결망을 보는 것도 이와 비슷하다고 생각한다. 즉, 현실의 연결망을 바라보는, 첫 번째도 아닌 영 번째 어림^{zeroth-order approximation}이 바로 “척도 없음”의 관점이다. 이 어림이 정확히 성립한다고 주장하는 것이 결코 아니다.

//

실제의 연결망을 보다 더 근본적으로
이해하기 위해서는

척도 없음이라는 연구의 틀을 넘어서
무언가가 더 필요할 수도 있다는 제
안이 바로 논문[1]의 진정한 결론이
다.

//

물론 논문[1]의 저자 클로셋은 물리학자들의 위의 설명에 만족하지는 않는 것으로 보인다. 클로셋의 주장을, 돌멩이를 떨어뜨렸더니 자유낙하에 비슷하게라도 떨어지는 돌멩이는 거의 없고, 심지어 떨어지지 않는 돌멩이가 더 많았다는 주장에 비유할 수 있겠다. 관찰된 상황은 같아도, 필자를 포함한 물리학자들은 정확한 자유낙하가 가능하지 않은 현실의 실험 상황을 탓하고, 논문[1]의 저자들은 어쨌든 자유낙하하지 않았으니 자유낙하라는 가정을 처음부터 의심해야 한다고 주장하는 셈이다.

필자의 생각도 홀메의 논문[2] 제목 “Rare and everywhere”와 비슷하다. 척도 없는 연결망은 도처에 있지만 척도 없음을 정말로 확인하는 것은 무척 어렵다. 보기 어렵다고 해서, 정말로 드물다는 결론을 곧바로 내릴 수는 없다. 그럼에도 불구하고, 논문[1]의 의미와 가치는 상당하다. 척도 없음이라는 영 번째 어렵만으로는 현실의 연결망을 거의 설명할 수 없다는 뼈아픈 지적이다. 자유낙하로 어렵해 돌맹이의 떨어짐을 대충 이해한 다음에 할 일은, 공기의 저항과 코리올리의 효과 등을 이해해 결과를 좀 더 정교하게 다듬는 일이다. 혹은, 점성이 아주 큰 유체 안에서 움직이는 아주 작은 물체라면 자유낙하라는 틀은 어쩌면 말도 안 되는 것일 수도 있겠다. 실제의 연결망을 보다 더 근본적으로 이해하기 위해서는 척도 없음이라는 연구의 틀을 넘어서는 무언가가 더 필요할 수도 있다는 제안이 바로 논문[1]의 진정한 결론이다. 연결망 다수에 대한 통계적 설명을 들었으니, 이제 다음 할 일은 하나하나에 대한 더 자세한 구체적 이해다. 두 논문 [1][2]의 결론은 둘 다 맞다. 척도 없는 연결망은 정말 드물다. 하지만 어디나 있다. 어디나 있을 것으로 믿어지지만 왜 이토록 드물게 발견되는 지, 우리는 아직 속속들이 그 답을 알지 못한다. 아직 갈 길이 남아있다.

참고문헌

1. A. D. Broido and A. Clauset, *Scale-free networks are rare*, Nat. Commun. 10, 1017 (2019).
2. P. Holme, *Rare and everywhere: Perspective on scale-free networks*, Nat. Commun. 10, 1016 (2019).
3. A. Clauset, C. Shalizi, and M. E. J. Newman, *Power-law distributions in empirical data*, SIAM Rev. 51, 661 (2009).
4. E. Klarreich, *Scant Evidence of Power Law Found in Real World Networks*, Quanta Magazine (Feb. 15, 2018).