전염확산 네트워크 연구

DOI: 10.3938/PhiT.24.052

고 광 일

Epidemics on Complex Networks

Kwang-Il GOH

Since the early 20th century, epidemic models have been studied as a major subject in mathematical biology. Driven initially by correspondence of these models with well-known problems in statistical physics such as percolation and more recently by their relevance in complex network theory, statistical physicists have also been actively studying epidemics. In this article, a brief overview of statistical physics studies on epidemic models, as well as current perspectives on network epidemics, will be given.

들어가는 글

전염병 유행은 인류의 역사와 함께해왔다. 특히 인류의 삶의 양식의 급격한 변화는 새로운 전염병의 유행을 수반해왔다. 선사시대 야생 동물의 가축화에 따른 다양한 동물원성감염증(zoonosis)의 유행이나 근대 도시화가 촉진시킨 흑사병의 유행 등이 대표적인 예이다. 최근 몇 년간 줄지어 창궐하고 있는 중증급성호흡기증후군(SARS), 조류독감(avian flu), 에볼라 바이러스(Ebola virus), 그리고 올해 우리나라를 떠들썩하게 달구었던 중동호흡기증후군(MERS) 등 다양한 전염병의 유행은 전염병이 인류의 삶에 끼치는 광범위한 경제적 사회적 영향을 우리 모두 몸소 체험하게 하고 있다. 이러한 연유로 그리스 시대 이래로 전염병의 확산 및 유행의 원인과 과정, 결과에 대한 이해는 중요한 과학적 문제였다. 본 소고에서는 전염병 확산 현상의 수리 모형에 대한 통계물리학 및 복잡계 네트워크 연구를 개괄한다. 관련 내용은 <물리학과 첨단기술> 2007년 10월호의 특집 기사 중 하나인 <복잡계에서의 전염확산 현상>^[1]과

저자약력

고광일 교수는 서울대학교 이학박사(2004년)로써 서울대학교 이론물리연구소 센터와 미국 University of Notre Dame 박사후연구원을 거쳐, 2007년부터 고려대학교 물리학과 교수로 재직 중이다.(kgoh@korea.ac.kr)

접치는 부분이 있음을 알려준다.

전염병 확산 현상에 대한 수리 모형 연구는 천연두를 대상으로 한 1766년 베르누이(Daniel Bernoulli)의 연구^[2]가 그 시초로 알려져 있으나, 현대적인 시작은 1927년, 1932년, 그리고 1933년 영국왕립학회지에 3편에 걸쳐 출판된 커맥과 맥켄드릭의 논문 <A Contribution to the Mathematical Theory of Epidemic>으로 본다.^[3] 이 논문에서 커맥과 맥켄드릭은 향후 전염병 확산 수리 모형의 기본이 되는 구획화(compartmental) 모형 체계를 제시한다. 이 모형 체계에서는 질병의 전염 상태에 따라 인구 집단을 여러 구획(compartment) 또는 군(群, class)으로 나누고 이 구획들 간의 상태 전이를 모형화한다. 가장 기본적인 SIR 모형의 경우 인구 집단은 질병에 걸릴 수 있는 '민감군(susceptible, S)', 질병에 걸려 전염시킬 수 있는 '전염군(infected, I)', 그리고 질병이 완전히 낫거나 사망하여 더 이상 전염 과정에 참여하지 않는 '제거군(removed, R)'의 세 가지로 나뉜다(각각의 구획의 영문 머리글자를 따라서 SIR 모형이라 부른다). 모형화하고자 하는 전염병의 특성 및 모형화의 세밀함 정도에 따라 다른 군을 도입할 수도 있다. 예를 들면, 독감(influenza)에는 질병에 노출되어 감염되었으나 아직 전염성이 없는 '노출군(exposed, E)'을 추가한 SEIR 모형을 많이 이용한다. 커맥과 맥켄드릭 이래로 거의 백년간 전염병 확산 수리 모형 연구는 수리 생물학의 주된 주제로 활발히 연구되어 다양한 변종 모형과 분석 방법이 개발되었다.^[4]

전염확산 모형과 통계물리학

전염확산 수리 모형의 기본 모형인 SIR 모형의 개체는 $S \rightarrow$

REFERENCES

- [1] J. D. Noh and S.-H. Yook, Phys. High Technol. **16**(10), 22 (2007).
- [2] D. Bernoulli, Mem. Math. Phys. Acad. Roy. Sci. Paris, 1 (1766); reprinted in Rev. Med. Virol. **14**, 275 (2004).
- [3] W. O. Kermack and A.G. McKendrick, Proc. Royal Soc. Lond. Ser. A **115**, 700 (1927); **138**, 55 (1932); **141**, 94 (1933); reprinted in Bull. Math. Biol. **53**, 33 (1991); **53**, 57 (1991); **53**, 89 (1991).
- [4] R. M. Anderson and R. M. May, *Infectious Diseases in Humans* (Oxford University Press, 1992).

$I \rightarrow R$ 의 과정을 따라 변한다. 감염군의 개체가 민감군의 개체와 접촉하면 β 의 비율(rate)로 전염되고($S \rightarrow I$), 감염군의 개체는 감염된 후 τ 의 시간이 지나면 회복되거나 죽게 되는($I \rightarrow R$) 간단한 전염확산 과정을 생각하자. 하나의 감염군 개체가 회복하기 전에 민감군 개체와 접촉하여 전염에 성공할 확률을 전파도(transmissibility) T 라고 부르고, 위 과정의 경우 이 값은 $T = 1 - \lim_{\Delta t \rightarrow 0} (1 - \beta \Delta t)^{1/\Delta t} = 1 - e^{-\beta \tau}$ 가 된다. 즉, 하나의 감염군 개체는 접촉한 민감군 개체를 T 의 확률로 전염시키고, 이렇게 전염된 감염군 개체가 접촉한 또 다른 민감군 개체를 T 의 확률로 전염시키는 과정이 계속되며 전염 확산이 진행된다. 전염병의 아웃브레이크(outbreak)는 전염 확산이 성공적으로 진행되어 인구 전반이 전염되는 상황을 일컫는다. 수학적으로는 전염확산이 끝없이 진행되어 무한한 인구의 일정 비율 이상이 전염될 확률이 확률을 아웃브레이크 확률 P 라고 하자)이 0보다 크게 되는 것을 뜻한다. 이러한 관점에서 전염병의 아웃브레이크는 통계역학계에서 나타나는 고비 현상을 보인다. 전파도가 임계값보다 작을 경우($T < T_c$) 아웃브레이크는 발생하지 않고($P = 0$), 전파도가 임계값을 넘어서야만($T > T_c$) 아웃브레이크가 발생한다($P > 0$). SIR 모형에서는 전염확산 과정의 결과 민감군(S)과 제거군(R)만이 남게 된다. 이 중 제거군이 질병에 감염되었던 개체들이고, 따라서 전염확산 과정 진행 후 제거군의 비율이 전염병 아웃브레이크의 크기 S 가 된다.

SIR 모형의 아웃브레이크 확률 P , 아웃브레이크 크기 S 는 다음과 같이 통계물리학의 주요 문제 중 하나인 스미기(percolation) 문제로 매핑(mapping)할 수 있다.^[5] 개체들 사이의 접촉 양상을 그래프로 표현한 네트워크를 생각하자. 즉, 개체 i 와 j 가 접촉하는 것을 i 와 j 를 잇는 링크로 표현한 네트워크이다. 이 네트워크는 격자 구조를 포함한 임의의 구조를 가질 수 있다. 감염군 개체 i 가 민감군 개체 j 를 전염시키는 일은 확률 T 로 일어나므로, 전염 과정을 표현하는 부분네트워크는 원래 네트워크의 링크가 확률 T 로 차있는(occupied) 본드 스미기(bond percolation) 클러스터가 된다. 이러한 매핑에 의해 아웃브레이크 확률 P 는 임의의 개체(감염군)가 본드 스미기의 거대 송이(giant cluster)에 속할 확률에, 아웃브레이크 크기 S 는 거대 송이의 크기(비율)에 대응되게 된다. 간단한 방향성 없는(undirected) 네트워크의 경우 두 값은 같다.^[6] 일반적으로 전파도에 방향성이 있는 경우에는 좀 더 정교한 매핑을 통해 아웃브레이크 확률 P 와 아웃브레이크 크기 S 가 각각 해당 방향성 있는(directed) 네트워크의 giant in-component와 giant out-component의 크기에 대응되고, 두 값은 일반적으로 같지 않다.^[7] 스미기 문제로의 매핑은 전염 확산의 결과인 아웃브레이크에 대한 정보를 주지만 동역학적인 정보를 주지 않는다.

전염 확산의 동역학 연구를 위해서도 속도식(rate equation), 몬테카를로 시뮬레이션 등 통계물리학의 도구가 널리 이용된다.

SIR 모형만큼이나 널리 연구되는 변종 모형이 소위 SIS 모형이다. SIS 모형에서 감염군의 개체는 감염된 후 시간 τ 가 흐른 뒤 다시 민감군으로 변환된다($I \rightarrow S$). 따라서 감염 전과 과정은 개체가 감염군과 민감군 상태를 반복적으로 순환하며 진행된다. 인구 전체가 민감군이 되는 경우 더 이상 전염 과정이 진행되지 못하는 동역학적으로 언(frozen) 상태가 되고 이를 흡수 상태(absorbing state)라고 부른다. 반면에 SIR 모형과 달리 무한한 인구에서 감염군의 비율이 0보다 큰 일정한 값으로 유지되며 질병 확산이 영원히 지속되는 정상상태가 추가로 존재할 수 있고, 이를 활성 상태(active state)라고 부른다. SIS 모형도 전파도(또는 전파율 $\lambda = \beta \tau$)에 따라 흡수 상태와 활성 상태 사이의 비평형 상전이상을 보인다.^[8]

SIS 모형은 격자구조 위의 확률 과정 모형 중 하나인 contact process^[9]와 매우 유사하다. 두 모형의 차이점은 SIS 모형에서는 감염군 개체가 단위 시간당 모든 연결된 개체를 β 의 비율로 전염시키는 반면, contact process에서는 감염된 상태인 격자점이 단위 시간당 하나의 이웃 격자점을 전염시킨다는 점이다. 격자 구조에서 이 차이는 단순히 시간 척도의 차이에 지나지 않는다. d 차원 격자에서의 contact process는 $(d+1)$ 차원의 방향성 스미기(directed percolation)로 매핑된다. 방향성 스미기의 고비 현상은 비평형 상전이의 가장 주요한 보편성 부류(universality class)를 형성하여 방대한 관련 연구가 진행되어 왔다.^[8]

전염확산 모형과 복잡계 네트워크

앞서 개괄한 것과 같이 전염 확산의 기본 모형은 통계물리학의 주요 모형인 스미기 문제와 다각도로 연관이 되며 이를 통해 직간접적으로 물리학의 관심 대상이었다. 그러나 여전히 전염확산 연구는 전통적인 수리 생물학의 영역이었고 물리학의 기여는 제한적이었다. 이러한 상황은 복잡계 네트워크의 등장과 함께 급변하여 지난 15년간 통계물리학은 전염확산 모형 연구에 근본적인 변화를 주도해 왔다. 복잡계 네트워크 연구에

REFERENCES

- [5] P. Grassberger, Math. Biosci. **63**, 157 (1983).
- [6] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
- [7] E. Kenah and J. M. Robins, Phys. Rev. E **76**, 036113 (2007).
- [8] M. Henkel, H. Hinrichsen and S. Lübeck, *Non-equilibrium Phase Transition: Absorbing Phase Transitions* (Springer-Verlag, 2008).
- [9] T. E. Harris, Ann. Prob. **2**, 969 (1974).

통계물리학이 주요한 기여를 해왔음은 주지의 사실이다. 전염 확산 모형 연구는 복잡계 네트워크 연구의 초창기에서부터 주목할 만한 결과를 얻어내며 주요 연구 주제로 자리매김해왔다.^[10]

복잡계 네트워크에서 보이는 전염확산 과정의 특이점은 대부분의 경우 복잡계 네트워크에 내재하는 연결선수(degree, 한 노드에 연결된 링크의 수)의 불균질성에서 유래한다. 이를테면, 앞서 제시했던 SIS 모형과 contact process의 차이는 연결선수가 균일한 격자 구조에서는 중요한 역할을 하지 않지만, 연결선수(degree)가 불균일한 척도없는 네트워크(scale-free network)의 경우 중요한 차이를 낳는다. 척도없는 네트워크란 연결선수 분포가 $p(k) \sim k^{-\gamma}$ 와 같이 멱함수를 따르는 네트워크로서, 매우 큰 연결선수를 가진 소위 '허브'가 존재하는 네트워크이다. 전염확산에서 이러한 연결선수 불균일성과 허브의 존재의 중요성을 각인시킨 최초의 결과는 평균장 이론기반 분석을 통해 척도없는 네트워크에서 SIS 모형의 임계 확산율 λ_c 이 $\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$ 이 되어 연결선수 분포의 지수값이 $\gamma \leq 3$ 일 때 $\lambda_c \rightarrow 0$ 가 됨을 주장한 것이다. 이는 연결선수의 비균일성이 충분히 크면 임의의 0이 아닌 확산율에 대해 전염병의 확산이 가능하다는 것을 의미한다.^[11] SIR 모형에 대해서도 비슷한 결론이 얻어졌다.^[6] 이러한 일련의 초기 연구를 통해 척도없는 네트워크에서 전염확산 과정은 매우 큰 연결선수를 가진 허브의 역할에 크게 의존하며 이러한 허브들이 전염확산을 매우 강력하게 촉진시킨다는 것이 상식이 되었다. 이와 관련하여 효율적인 전염확산에 기여하는 노드를 특징짓는 연결선수보다 더 나은 특성을 찾는 소위 효율적 확산자(efficient spreader 또는 influential spreader) 연구가 하나의 주제를 형성하고 있다.

척도 없는 네트워크의 허브의 존재는 전염병 확산을 효과적으로 예방하는데도 이용될 수 있다. 허브가 전염병 확산에 큰 기여를 하므로, 허브의 활동을 막아 확산을 효율적으로 예방할 수 있는 기회가 생기는 것이다. 문제는 허브를 어떻게 찾는가이다. 이는 네트워크 구조를 완벽하게 알고 있는 경우가 아니라면 일반적으로 쉽지 않은 문제이다. 이를테면 성적 접촉에 의해 전파되는 질병 전파 네트워크의 허브를 찾는 문제를 생각해 보면 될 것이다. 이 문제의 가장 유명한 해결책으로 소위 '지인 면역(acquaintance immunization)'이라 불리는 방법이 소개되었다.^[12] 이 방법은 네트워크의 전체 구조를 모르는 상태에서 연결선수가 많은 허브를 찾기 위해 먼저 임의의 노드를 무작위로 찾고, 그 노드에 연결된 노드 중 하나를 다시 무작위로 골라 선택하여 면역시킨다. 임의로 고른 노드의 평균 연결선수는 $\sum_k kp(k) = \langle k \rangle$ 이지만, 임의로 고른 노드에 연결된 노드는 임의로 고른 링크에 연결된 노드와 같으므로 특정 노드

가 선택될 확률이 그 노드의 연결선수에 비례하여 평균 연결선수가 $\sum_k k \frac{kp(k)}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$ 가 된다. 따라서 이 방법을 통해 연결선수가 큰 허브를 확률적으로 더 선택할 수 있게 된다.

전염확산 네트워크 연구 최신 동향

척도 없는 네트워크에서의 전염확산의 특이성에 주목한 초기 연구 이래로 이 분야는 복잡계 네트워크 동역학 연구의 주된 연구 주제가 되고 있다. 평균장 이론 분석이 불충분한 SIS 모형의 경우 임계 확산율에 대한 더욱 자세한 연구가 활발히 진행되고 있으나 아직까지 최종 해답이 내려지지 않은 상황이다. SIR 모형 및 다양한 변종 모형에 대해서도 더욱 정교한 스미기 모형으로의 매핑을 통한 연구가 활발히 진행되고 있다.^[10]

복잡계 네트워크 이론의 발전에 발맞춰 새로운 구조 및 동역학적 개념이 전염확산 연구에 적용되고 있다. 그 중 최근 주목받고 있는 것이 접촉 네트워크 자체의 시간적 변화이다. 네트워크의 동적 변화는 크게 두 수준에서 다루어질 수 있다. 첫째로 접촉 네트워크의 구조 자체는 변하지 않으나 각 링크의 활동성(activity)이 시간적 변화를 겪는 경우이다. 링크의 변화 시간 척도가 접촉 동역학의 시간 척도보다 충분히 길 때 유효한 방법이다. 이 경우 중요한 물리량은 링크의 시간적 활동성을 기술하는 대기 시간 분포(waiting time distribution) $p(\tau)$ 이다. 대기 시간이란 주어진 링크의 연이은 활동 시간 사이의 간격을 말한다. 특히 대기 시간 분포가 $p(\tau) \sim \tau^{-\alpha}$ 와 같이 멱함수 꼴이 되면 매우 긴 대기 시간 동안 링크가 휴지기로 전염 확산에 참여하지 않게 된다. 이때 특히 대기 시간 분포의 지수가 $\alpha \rightarrow 2$ 로 접근하면 SIR 모형의 임계 확산율이 $\lambda_c \rightarrow \infty$ 로 발산하게 되어 전염확산이 극단적으로 억제됨이 보고되었다.^[13] 두 번째는 접촉 네트워크의 구조 자체가 시간에 대해 변화하는 경우이다. 이러한 네트워크 구조를 시간적 네트워크(temporal network)라 부른다. 링크의 변화 시간 척도가 접촉 동역학의 시간 척도와 엇비슷할 때 유효한 방법이다. 이 경우는 활동성 분포(activity distribution) $p(a)$ 가 중요한 역할을 한다. 노드 i 의 활동성 a_i 는 단위시간당 이 노드가 활동적일(전염확

REFERENCES

- [10] R. Pastor-Satorras, C. Castellano, P. Van Mieghem and A. Vespignani, Rev. Mod. Phys. **87**, 925 (2015).
- [11] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
- [12] R. Cohen, S. Havlin and D. ben-Avraham, Phys. Rev. Lett. **91**, 247901 (2003).
- [13] B. Min, K.-I. Goh and I.-M. Kim, EPL **103**, 50002 (2013).

산에 참여할) 확률로 정의된다. 이러한 소위 활동성 기반(activity-driven) 전염 확산 연구의 예로, 무작위 temporal networks에서 SIS 모형의 임계 확산율이 $\lambda_c \sim \frac{1}{\langle a \rangle + \sqrt{\langle a^2 \rangle}}$ 와 같이 활동성 분포에 의존함이 보고되었다.^[14] 또한 temporal network에서는 네트워크의 시간적 변화가 유도하는 접촉 동역학의 인과성이 올바르게 고려되어야 한다.

복잡계 네트워크 이론의 또 하나의 최신 주제는 네트워크의 다층위성(multiplexity)이다. 사람들 사이의 사회 연결망은 다층 위 네트워크의 주요한 예로 친구 네트워크, 가족 네트워크, 협업 네트워크 등 다양한 층위의 사회연결망 레이어가 전염 확산에 기여한다. 각각의 레이어는 매우 다른 동역학적 특성을 가지며 시공간적으로 약하게 결합되어 있다. 이러한 네트워크의 다층위성이 전염확산에 미치는 영향이 다각도로 연구되고 있다. 그 예로 전염확산이 서로 다른 레이어를 통해 일어날 때의 교차 비용(crossing cost)의 효과에 대한 연구,^[15] 전염병 전파 네트워크 레이어와 전염병 확산 정보가 퍼지는 사회연결망의 경쟁으로 인한 효과에 대한 연구^[16] 등이 보고되었다.

전염확산 수리 모형 연구의 궁극적인 목표는 다양한 단순 모형 연구를 통한 기본 원리 이해를 넘어서 최대한 현실적인 모형을 통한 전염확산 모니터링 시스템 구축과 정량적인 아웃브레이크 예보를 실현하는 것이다. 이를 목표로 전 세계 여러 연구 그룹들이 대용량 정보 기반의 전염 확산 예측 모형을

개발하고 있다. 휴대폰 위치 정보 기반의 국지적 개인 이동 정보에서부터 항공운항 정보 기반의 국가 간 이동 정보를 반영하여 구축된 이러한 상세 모형들은 중증급성호흡기증후군, 신종플루, 에볼라 바이러스 등의 예를 거치며 그 성능이 나날이 향상되고 있다. 대표적인 예로 FLUOUTLOOK(<http://www.fluoutlook.org>) 등의 온라인 예측 플랫폼이 운영되고 있다.

맺는 글

전염병 확산과 창궐의 경제적 사회적 파급효과가 지속되는 한 전염확산 연구에 대한 관심은 계속될 것이다. 변화하는 사회와 함께 전염확산 연구도 끊임없이 진화하고 있고, 앞으로도 그럴 것이다. 본 소고에서는 통계물리학의 관점에서 이제까지 진행된 전염확산 모형 연구와 전염확산 네트워크 연구의 최신 동향을 간략하게 개괄하였다. 이 방대한 연구 분야에 대한 더욱 자세한 내용은 앞서 제시한 참고 문헌^[1,4,10]을 참조하시길 당부 드린다.

REFERENCES

- [14] N. Perra, B. Goncalvez, R. Pastor-Satorras and A. Vespignani, Sci. Rep. 2, 469 (2012).
- [15] B. Min and K.-I. Goh, arXiv:1307.2967.
- [16] C. Granell, S. Gomez and A. Arenas, Phys. Rev. Lett. 111, 128701 (2013).