

복잡계 네트워크의 구조적 특징과 동역학 현상

강병남* · 고광일 · 이덕선 · 김두철

서울대학교 자연과학대학 물리학과 NS50, 서울 151-747

(2004년 1월 27일 받음)

최근 들어 복잡계 네트워크에 대한 연구가 학제간 연구 주제로 많은 관심을 불러 일으키고 있다. 이 연구는 사회, 경제, 생명체, 인터넷 등과 같은 복잡계에서 일어나는 다양한 현상을 그래프 이론을 빌려 설명하고자 하는 것이다. 이와 같은 다양한 계에서 나타나는 복잡계 네트워크의 공통된 특징은 임의의 한 노드에 연결된 연결선수가 매우 불균일하며 그것의 분포가 임계현상에서 볼 수 있는 바와 같은 멱함수 법칙을 보인다는 것이다. 본 총설논문에서는 최근 3-4년간 폭발적으로 이루어진 복잡계 네트워크에 대한 주요한 연구결과, 즉 실재하는 복잡계 네트워크의 구조적 특징과 이에 대한 모형, 네트워크 구조의 시간적 진화현상, 또 네트워크 위에서 일어나는 동역학 현상에 대한 기본적인 내용을 소개하고자 한다.

PACS numbers: 64.60.-i

Keywords: 복잡계, 네트워크, 멱함수법칙

I. 들어가기

1960년대 말 하버드 대학의 사회학자인 밀그램이 실험적으로 보인 “지구상의 사람들은 5 단계만 거치면 거의 다 알게 된다”고 하는 소위 여섯단계 분리(Six degrees of separation) [1]가 알려진 이후, 좁은 세상 (small world)이라는 단어는 우리들 생활 속에 깊이 자리를 잡고 있다. 사회학에서 그래프 이론을 이용하여 사람과 사람과의 관계를 이해하고자 하는 사회연결망에 대한 연구는 1960년대 이후 많은 관심을 불러 일으키고 있다. 이러한 연구는 제 III 절에서 소개할 에르도쉬와 레이니(ER)의 무작위 네트워크(random network)에 대한 논문 [2]의 영향에서 비롯되었다고 할 수 있다. 이 ER의 논문은 현대 그래프 이론의 효시가 되었다. 그래프는 여러 점들과 이러한 점들을 연결하는 연결선으로 구성되어 있는데, 점과 연결선을 각각 노드(node)와 링크(link)라고 부른다. 최근에는 인터넷의 영향으로 그래프라는 용어 대신 네트워크라는 용어가 친숙해 있기 때문에 본 논문에서도 그래프라는 용어대신 앞으로 네트워크 또는 연결망이라는 용어를 사용하기로 하겠다.

네트워크 이론은 사회학에서만 유용하게 쓰이는 것이 아니고, 생물학, 경제학, 컴퓨터공학 등 여러 분야에서 유용하게 쓰인다. 따라서, 네트워크의 구조적 성질을 살

펴보는 것과 네트워크 위에서 일어나는 다양한 현상을 연구하는 것은 매우 흥미롭고 중요한 문제라고 할 수 있다. 사회연결망에서 노드는 사람에게 해당되고 링크는 사람과 사람과의 관계에 해당된다. 아주 단순하게 두 사람이 서로 아는 사이이면 링크가 있고, 모르면 링크가 없는 경우에서 출발하자. 물론 서로 얼마나 잘 아는 사이인가라는 문제는 좀 더 복잡한 문제이다. 이렇게 사람과 사람들 간의 관계를 네트워크로 그려놓고 여기서 드러나는 성질, 예를 들어 집단화 등에 대하여 이해하는 것은 매우 유용한 방법이라 할 수 있다. 이 방법은 조직관리 등에 유용하게 사용할 수 있기 때문에 사회학에서 관심을 가지고 연구하고 있다 [3]. 생물학에서도 최근 네트워크 이론을 빌려 생물체에서 발견되는 여러 가지 생리현상을 이해하고자 하는 연구가 경쟁적으로 이루어지고 있다. 예를 들어 단백질 상호작용 네트워크 (protein interaction network)를 생각해 보자. 이 네트워크에서는 노드가 단백질을 의미하고 두 단백질이 서로 물리적으로 상호작용이 가능하면 서로간의 연결선이 존재한다고 하는 방법으로 네트워크가 만들어진다. 이러한 네트워크를 구성하고 나면 생화학반응 또는 생리현상을 제어할 수 있는 방법을 찾는 데 편리하고, 유사한 단백질들의 공동 역할을 전체의 구조속에서 쉽게 파악할 수 있기 때문에 많은 연구가 수행되고 있는 것이다. 특히 유전자 배열이 밝혀지면서 유전자가 만들어 내는 단백질의 상호작용에 많은 관심을 갖는 것은 당연한 일이고, 네트워크

*E-mail: kahng@phya.snu.ac.kr

크를 이용한 연구는 이 분야에서 폭발적으로 수행되고 있다고 할 수 있다 [4]. 오늘날 흔히 생각할 수 있는 인터넷도 정보 네트워크의 전형적인 예이다. 여기서 노드는 라우터 또는 자동 제어시스템(autonomous system)에 해당되고, 두 라우터가 케이블로 연결되어 있으면 연결선이 있는 것으로 생각할 수 있다 [5].

최근 네트워크 연구가 ER 이후 많은 관심을 다시 받게 된 것에 큰 역할을 한 것은 왓츠와 스트로가츠(WS)가 발표한 논문 [6]이었다. WS는 영화배우 네트워크, 선충(C. elegans)의 신경망, 전력망에서 결집계수(clustering coefficient)라는 수학적 양을 측정하였다. 그들은 이러한 실재하는 네트워크에서의 성질이 ER의 무작위 네트워크 모형이 주는 결집계수와 많은 차이가 있음을 알게 되었고 이를 설명하기 위해 ‘좁은세상 네트워크’라고 부르는 모형을 제시하였다. 여기서 결집계수라고 하는 것은 뒤에 자세히 소개하겠지만 네트워크에서 세 개의 노드를 잇는 삼각형이 얼마나 많이 존재하는가를 나타내는 수학적 양이다. 이 모형은 제 V 절에서 자세히 소개하겠지만 규칙적인 유클리디안 네트워크를 약간 변형시킨 것으로, 실제 존재하는 여러 네트워크의 결집계수와 비슷한 값을 준다는 측면에서 좋은 모형이 되었지만 네트워크의 다른 구조적 특징을 줄 수 없어 크게 호응을 받지 못하게 되었다.

1998년 WS의 논문이 발표된 시점에 사람들은 인터넷의 영향으로 월드-와이드 웹의 구조에 대하여 관심을 가지게 되었다. 알버트, 정, 바라바시 [7]와 후버만, 아다믹 [8]은 월드-와이드 웹 구조에 관심을 가지고 연구하였다. 월드-와이드 웹이란 하이퍼 링크로 연결된 서로 다른 웹-페이지들이 노드가 되는 네트워크를 의미한다. 이들의 연구 결과는 어떤 한 노드에 연결된 연결선수는 노드마다 매우 불균일하여, 이들의 분포함수는 멱함수(power law)를 따른다는 것이었다. 여기서 연결선수라는 것은 어떤 웹-페이지 A 에서 하이퍼 링크된 페이지의 개수를 의미하는 나가는-연결선수(out-degree)와 A 라는 웹-페이지는 모르지만 다른 웹-페이지들에서 A 라는 웹-페이지로 하이퍼 링크하는 경우 A 라는 웹-페이지로 들어오는, 들어오는-연결선수(in-degree)가 있을 수 있다. 연구 결과는 나가는-연결선수와 들어오는-연결선수 분포 함수 모두 멱함수의 꼴을 따라 행동한다는 것이었다. 이의 결과는 매우 흥미로운 것이었다. 멱함수 법칙은 강자성체가 상자성체로 상전이하는 임계현상에서 나타나는 것처럼, 어떤 특정한 척도가 없다는 것을 의미하기 때문에 이러한 성질을 갖는 네트워크를 척도없는-네트워크라고 부른다 [9].

월드-와이드 웹에 대한 논문이 발표된 이후 최근 3-4년 동안 복잡계 네트워크에 대한 논문이 폭발적으로 발표되고 있다. 이에 대한 배경은 복잡계가 21세기의 주요한 연구 대상으로 떠오르고 있기 때문이라고 생각된다. ‘복잡계란 무엇인가’라는 질문에 대한 명확한 정의는 아직 정립되어 있지는 않지만, 넓게 보아서 복잡계란 많은 구성원들이 존재하고 서로간의 상호작용을 통하여 만들어 내는 거시적 현상이 아주 다양한 시스템이라고 할 수 있다 [11-13]. 수억 개의 웹-페이지들이 서로를 하이퍼링크하면서 만들어 내는 연결선수 분포함수가 멱함수 분포를 이루고 있다는 것이 복잡계에 어떤 질서가 존재함을 보여주는 한 예라 할 수 있다. 앞에서 언급한 바와 같이 연결선수 분포가 멱함수를 따른다는 것은 종래의 임계현상에서 보여주는 열역학적 물리량들이 멱함수 분포의 특징을 보여주는 것과 그 정신이 비슷하다. 예를 들어 강자성체에서 상자성체로의 상전이 현상을 기술하는 이징모형에서 임계온도에서 같은 방향의 스핀들이 만들어 내는 스핀도메인의 크기 분포가 멱함수를 따른다는 것은 잘 알려진 사실이다. 여기서 보편성이란 지수나 그 외의 물리량들이 계의 미시적이고 구체적인 내용에는 의존하지 않고 보편적인 값을 취한다는 개념이다. 이러한 스핀계에서 멱함수의 지수는 보편성을 지니고 있는 양이다. 이와 같이 스핀계에서는 보편성이 존재하여 흥미롭지만 복잡계 네트워크에서는 보편성과 축척법칙이 어떻게 나타날지에 대한 명확한 이론은 아직 정립되어 있지 않은 상태이다. 그러나 스핀계에서 연구되었던 방법을 이용하여 다양한 복잡계 네트워크에서 보편성을 찾는다는 것은 매우 중요한 일이므로 이러한 방향으로의 연구가 지속적으로 수행될 것으로 기대한다.

본 총설논문에서는 최근에 폭발적으로 수행되고 있는 복잡계 네트워크, 특히 척도없는-네트워크의 구조적인 특징을 살펴볼 것이며, 또한 척도없는-네트워크 위에서 일어나는 동역학적 문제들도 살펴보기로 하겠다. 이 분야는 짧은 시간동안에 많은 논문들이 발표되고 있기 때문에 모두 다 다룰 수는 없고, 본 연구팀의 연구 관심사를 중심으로 선택된 내용임을 밝힌다. 일반적인 내용에 대하여서는 최근 비전공자, 또는 전공자들을 위한 책과 총설논문들이 많이 나오고 있기 때문에 그것들을 참고해 주기 바란다. 우선 비전공자를 위한 책으로는 바라바시 [14], 뷰캐넌 [15], 왓츠 [16], 후버만 [5], 카프만 [17]이 쓴 책이 있다. 또 전공자들을 위한 책으로는 분홀트와 슈스터가 편저한 책 [18] 이 있으며, 도로고체프와 멘데스 [19], 볼로바스 [20], 발디 [21] 등이 쓴 책이 있다. 총설논문으로는 알버트와 바라바시가 쓴 논문 [22], 도로고체프와 멘데스가 쓴 논문 [23] 과 뉴우만이 쓴 논문 [24] 이 대

표적인 것이므로 이를 참고하기 바란다. 또한 앞으로도 최근에 연구된 논문들을 포함하여 편저 또는 저술한 책 및 논문들이 속속 출판될 예정이다.

II. 네트워크 분석에서 사용되는 물리량

이 절에서는 본 총설논문에서 공통적으로 사용하고 있는 물리량에 대하여 소개하고자 한다. 이러한 양들은 네트워크의 구조적 특징을 이해하는데 도움이 된다.

첫째, 어떤 계(system)를 생각하자. 이 계는 총 N 개의 노드와 L 개의 연결선으로 구성되어 있다고 하자. N 과 L 이 시간이 변하면서 증가하는 경우가 있는데 이 경우의 네트워크를 성장-네트워크(growing network)라고 한다. 이에 반하여 N 과 L 이 시간에 무관한 경우의 네트워크를 정적-네트워크(static network)라고 한다. 성장-네트워크에서 특별히 N 의 성장 속도보다 L 의 성장 속도가 빠른 경우를 가속적 성장-네트워크(accelerated growing network)라고 한다 [25]. 가속적 성장-네트워크는 성장-네트워크와는 다른 특성이 있다. 또, 연결선에 방향이 있는 경우와 방향이 없는 경우가 있는데, 전자의 경우를 방향성-네트워크(directed network), 후자의 경우를 비방향성-네트워크(undirected network)라고 한다.

둘째, 어느 한 노드 i 에 연결된 연결선수를 k_i 라 하자. 이를 그래프이론에서는 도수(degree)라 부르는데, 이 논문에서는 ‘연결선수’라 부르기로 한다. 각 노드들의 연결선수, $\{k_i\}$ ($i = 1 \dots N$), 로 구성되는 분포함수를 연결선수 분포함수 $P_d(k)$ 라 한다. 이는 주어진 하나의 네트워크에서는 $P_d(k)$ 는 연결선수가 k 인 노드의 수를 총 노드 수 N 으로 나눈 양으로 정의되고, 네트워크의 앙상블에 대하여서는 이의 앙상블 평균으로 정의된다. 이 분포함수가 멱함수를 따른다는 것은 k 가 클 때

$$P_d(k) \sim k^{-\gamma}, \quad (1)$$

을 의미한다. 여기서 지수 γ 를 연결선수 지수(degree exponent)라고 한다. 이 논문에서 분포함수는 항상 규격화되어 있다고 생각한다. 예를 들어 $\sum_k P_d(k) = 1$. 물리적으로 식 (1)은 지수함수보다 천천히 감소한다는 것을 의미하고, 이런 분포함수는 때로는 꼬리가 뚱뚱한(fat-tail이 있는) 분포함수라고 한다. 또한 식 (1)을 일반화하여,

$$P_d(k) \sim (k + k_0)^{-\gamma} \quad (2)$$

와 같은 분포함수를 생각할 수 있는데 이를 일반화된 Pareto 식이라고 부른다 [26-28]. 여기서 k_0 는 상수이다.

연결선수의 평균값 $\langle k \rangle = \sum_k k P_d(k)$ 은 $2L/N$ 이고 임의의 연결선을 택하였을 때 그 연결선의 한 쪽 끝의 노드의 연결선수가 k 일 확률은 $k P_d(k) / \langle k \rangle$ 이다.

셋째, 결집계수는 다음과 같이 정의된다. 어느 한 노드 i 를 잡았을 때 이 노드의 연결선수를 k_i 라 할때 k_i 개의 이웃 노드들끼리 얼마나 잘 연결되어 있는가를 나타내는 양이다. 즉

$$C_i = \frac{2E_i}{k_i(k_i - 1)}. \quad (3)$$

여기서 E_i 는 노드 i 의 k_i 의 이웃들 사이에 연결된 링크의 수를 의미하고, $k_i(k_i - 1)/2$ 는 그들간의 가능한 모든 연결선수이다. 한편 k 개의 연결선이 연결되어 있는 노드들의 평균 결집계수를 $C(k)$ 로 표시한다. 네트워크가 계층적 집단구조(hierarchical modular structure)를 가지는 경우에는

$$C(k) \sim k^{-\eta} \quad (4)$$

과 같이 멱함수로 감소하게 된다 [29]. 또, 모든 노드에 대한 평균 결집계수를 C 라 표시하는데, $C = (1/N) \sum_i C_i$ 이며, 이것은 $C = \sum_{k=1} C(k) P_d(k)$ 와 같다.

넷째, k 개의 연결선이 연결되어 있는 노드의 이웃 노드들의 평균 연결선수를 $\langle k_{nn} \rangle(k)$ 라 표시한다 [30]. 이 양은 하나의 연결선 양쪽에 붙어 있는 노드의 연결선수에 대한 서로의 상관관계를 이해하는데 유용하게 쓰인다. $\langle k_{nn} \rangle(k)$ 는 k 가 증가함에 따라 감소하는 경우와 증가하는 경우가 있다. 감소하는 경우는 평균적으로 연결선수가 많은 노드와 적은 노드가 서로 연결되어 있다는 것을 의미하고, 증가하는 경우는 연결선수가 많은 노드는 많은 노드와 연결되어 있고, 적은 노드는 적은 노드와 연결되어 있다는 것을 의미한다. 최근 인터넷과 같은 복잡계 네트워크나 연결선수 간의 상관관계가 있는 경우에는 위의 양이 멱함수에 따른다는 것이 밝혀졌다. 즉

$$\langle k_{nn} \rangle(k) \sim k^{-\nu}. \quad (5)$$

다섯째, 유사성 지표(assortativity index)는 다음과 같이 정의된다 [31].

$$r_k = \frac{\langle k_1 k_2 \rangle - \langle (k_1 + k_2)/2 \rangle^2}{\langle (k_1^2 + k_2^2)/2 \rangle - \langle (k_1 + k_2)/2 \rangle^2},$$

여기서 k_1 와 k_2 는 어떤 연결선의 양쪽에 붙어있는 노드의 연결선수에서 1 을 뺀 양이다. 이것은 다름 아닌 피어슨 상관관계 계수이다. 유사성 지표 r_k 이 양수이면 평균적으로 연결선수가 많은 노드는 많은 노드와, 적은 노드는 적은 노드와 연결된다는 것을 의미한다. 이에 반해 r_k

이 음수인 경우에는 연결선 수가 많은 노드와 적은 노드가 연결된다는 것을 의미한다. $r_k = 0$ 인 경우에는 아무런 경향이 없는 경우를 의미한다. $r_k > 0$ 인 경우는 $\langle k_{nn} \rangle$ 에서 정의한 $\nu < 0$ 인 경우에 해당되며, $r_k < 0$ 인 경우는 $\nu > 0$ 인 경우에 해당된다. $r_k > 0$ 인 네트워크를 유사 assortative 네트워크라 하고, $r_k < 0$ 인 네트워크를 비유사(dissortative) 네트워크라 하고, $r_k = 0$ 인 네트워크를 중립 네트워크라고 한다.

마지막으로 네트워크의 지름(diameter)이란 임의의 두 노드간의 평균 거리를 의미한다. 여기서 두 노드간의 거리란 한 노드에서 다른 노드로 가는데 거쳐야 하는 최소의 노드수를 의미한다.

III. 정적 무작위 네트워크

1. 여과군의 형성

1960년 ER은 간단한 네트워크의 진화현상을 살펴 보았다 [2]. 진화라는 용어는 초기부터 N 개의 노드가 존재하는 정적 네트워크에서 연결선 수 L 이 변함에 따라 네트워크가 어떻게 변하는가를 의미한다. ER이 제안한 무작위 네트워크(random network)에서는 N 개의 노드에서 임의로 2개의 노드를 무작위(random)적으로 뽑아 그 두 노드 사이에 연결선이 없으면 새로이 연결선을 잇고, 연결선이 있으면 다시 두개의 노드를 뽑아 연결선의 존재 여부를 판단하여 없으면 잇고, 있으면 또 다른 2개의 노드를 뽑고 하는 식으로 하여 L 개의 연결선을 연결한다. ER는 이와 같은 무작위 네트워크에서의 진화현상을 이해하기 위해 앙상블 개념을 이용한 확률적인 방법을 이용하였다. 시스템에 N 개의 노드가 존재하는 경우 가능한 두 노드간의 쌍은 $N(N-1)/2$ 개 있을 수 있다. 이때 두 노드 간에 연결선이 존재할 확률을 p 라 하면 이 시스템에 존재하는 연결선수의 기대치는

$$\langle L \rangle = pN(N-1)/2 \quad (6)$$

이 된다. 따라서 연결선이 존재하는 확률 p 를 조절함으로써, 시스템에 존재하는 연결선수의 기대치를 변하게 할 수 있다. 우리는 앞으로 p 를 조절하면서 네트워크가 어떻게 변하는가를 살펴보기로 하겠다. ER이 특히 관심을 가진 양은 p 가 증가함에 따라, 대부분의 노드들을 연결하는 송이(spanning cluster)가 언제 만들어질 것인가 하는 것이었다. 오늘날의 용어를 빌리면 여과 전이(percolation transition)가 언제 일어날 것인가 하는 것이었다. 즉 p 가 작을 때는 소수의 노드들만이 연결된 상태가 되어 있다

가, p_c 라는 값이 존재하여 $p > p_c$ 가 되면 계의 $\mathcal{O}(N)$ 노드들이 연결되는 큰 송이가 형성된다. 이때 p_c 를 임계 확률이라 부른다. ER는 해석적으로 $p_c(N) \simeq 1/N$ 임을 보였다. 이에 대한 유도 과정은 뒤에 척도없는-네트워크의 여과전이에 대한 논의를 참조하면 될 것이다. 이 임계 확률에서 존재하는 총 연결선수의 기대치는 $\langle L_c \rangle \simeq N/2$ 이다. 어느 한 노드에 연결된 평균 연결선수는 다음의 관계,

$$\langle k \rangle = 2\langle L \rangle / N, \quad (7)$$

로부터 $\langle k_c \rangle \simeq 1$ 이 된다. 즉 각 노드 당 평균 하나의 연결선이 연결되기만 하면 전체를 연결하는 여과송이가 만들어진다. 임계점 $p = p_c$ 에서 생기는 여과송이의 크기는 $\mathcal{O}(N^{2/3})$ 으로 알려져 있으며 임계점보다 큰 경우에는 즉 $p > p_c$ 에서는 그 크기가 $\mathcal{O}(N)$ 이 된다. 크기가 유한한 송이들의 평균크기 $\langle s \rangle$ 는 임계점에서 발산을 하는데 $\langle s \rangle \sim |p - p_c|^{-1}$ 로 발산한다. 또 임계점에서 s 라는 크기를 갖는 송이의 개수는 $n(s) \sim s^{-5/2}$ 에 따른다. 이상과 같은 특징을 가진 무작위 네트워크의 여과전이는 이차상전이의 성질을 보인다.

2. 연결선수 분포

각 노드 당 연결된 연결선수의 분포를 살펴보기로 하자. 역사적으로 돌이켜 보면 ER는 연결선수의 최소치와 최대치 만을 구했고, 헝가리의 수학자 볼로바스가 1981년에 무작위 네트워크에 대한 연결선수 분포함수를 구하였다 [32]. 어떤 계에 N 개의 노드가 있다고 하자. 어떤 한 노드는 $N-1$ 개의 다른 노드들과 연결이 가능하고 각각의 연결선은 p 라는 확률로 존재한다. 따라서 어느 한 노드에 k 개의 링크가 연결될 확률은 다음과 같이 주어진다.

$$P_d(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \quad (8)$$

N 이 매우 큰 경우 그러나 p 가 매우 작아 평균 $(N-1)p$ 이 유한한 경우에는 식 (8) 은 푸아송 분포, 즉

$$P_d(k) \simeq e^{-pN} \frac{(pN)^k}{k!} = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \quad (9)$$

이 된다. 따라서, 무작위 네트워크의 연결선수의 분포는 푸아송 함수를 따르게 된다.

3. 두 노드간의 평균거리

무작위 네트워크에서의 지름은 2001년 청과 루 [33]에 의해서 다음과 같이 구하여졌다. 무작위 네트워크에서 임계 확률 p_c 보다 약간 큰 확률로 연결선이 존재하여 전체 노드들이 서로 연결되어 있다고 하자. 즉 큰 송이에 전체 N 개의 노드가 포함되어 있다. 각 노드들은 평균적으로 $\langle k \rangle$ 개의 연결선이 연결되어 있다고 하자. 이때 한 노드에서 임의의 다른 노드로 가기 위해 평균적으로 D 개의 노드를 거쳐야 한다고 하자. 한 노드에서 출발하여 $\langle k \rangle$ 개의 이웃 노드 중에서 한 노드를 선택하여 그곳으로 이동하고 그곳에 있는 $\langle k \rangle$ 개의 노드 중에서도 임의의 한 이웃 노드를 골라가고 하는 식으로 D 걸음을 걸었다고 하자. D 걸음을 걸어 전체 N 개의 노드 중 어느 노드에도 도착할 수 있다고 한다면 D 걸음 후 어느 한 노드에 위치하여 있을 확률로부터 다음과 같은 식이 성립한다.

$$\frac{1}{\langle k \rangle^D} = \frac{1}{N}. \quad (10)$$

따라서 지름 $D = \ln N / \ln \langle k \rangle$ 이다. 무작위 네트워크에서 지름은 계의 총 노드 수 N 에 $\ln N$ 으로 비례하게 된다. 흥미롭게도 가속적 성장 네트워크에서는 즉 $L \sim N^{1+\theta}$ ($\theta > 0$) 인 경우에는 평균 연결선수가 $\langle k \rangle \sim N^\theta$ 이 되고, 따라서, 지름은 N 에 무관하게 된다.

4. 결집계수

각 연결선은 p 라는 확률을 가지고 존재하므로 k 라는 연결선수를 가지고 있는 어떤 노드의 이웃 노드들끼리 가질 수 있는 가능한 총 연결선수 $k(k-1)/2$ 중에서 존재하는 연결선수는 $pk(k-1)/2$ 이다. 따라서, 결집계수는 간단히 $C = p$ 이다.

IV. 성장 무작위 네트워크

ER 모형에서 생기는 여과전이 현상은 정적인 모형에서 즉 총 노드수는 고정되어 있고 연결선수가 증가하는 네트워크에서 여과송이가 언제 형성되는가에 대한 문제였다. 이때 생기는 상전이는 이차상전이에 따랐다. 한 편 노드의 수가 변하는 경우 즉 성장-네트워크에서의 여과전이는 코스터리츠-싸우리스 전이의 보편성 군에 따른다는 것이 밝혀졌다. 즉 무한차 상전이의 성질을 가짐이 밝혀졌다. 이에 대한 연구는 칼라웨이와 그의 공동연구

Table 1. Comparison of the diameter D and the clustering coefficient C of the actor network, the power grid, and the C. elegance neural network with those for corresponding random networks with the same numbers of nodes and links.

	D	D_r	C	C_r
Actor network	3.65	2.99	0.79	0.0027
Power grid	18.7	12.4	0.090	0.005
C. elegance	2.65	2.25	0.28	0.05

자 [34]에 의하여 수치적으로 처음 제안되었는데 이후 로고체프 등이 해석적으로 보였다 [35]. 그 모형은 다음과 같다. (i) 매 시간 새로운 노드가 태어나는 경우를 생각하자. 따라서 시간 t 가 지나간 후에 계에 존재하는 총 노드의 수는 t 이다. 이 새로운 노드는 태어날 때 자기 혼자있는 노드 상태로 존재한다. (ii) 매 시간 마다 b 개의 연결선이 기존의 노드들 사이에서 만들어 진다. 여기서 연결선은 임의의 두개의 노드를 선택하여 연결선이 없으면 연결선을 연결하는 방법으로 만들어 진다. 여기서 b 라는 것이 조절 변수가 되는데 일반적으로 b 는 자연수일 필요는 없다. 예를 들어 $b = 1/2$ 이면 2 단위시간이 지나면서 평균적으로 한개의 연결선이 만들어 진다고 생각하면 된다. 이 모형의 결과는 다음과 같다. (i) $b > b_c = 1/8$ 에 대하여 여과송이가 만들어지며 이 여과송이의 크기 $S(b)$ 는 $b < b_c$ 일때 0 이며, $b > b_c$ 일때,

$$S(b) = 0.59 \exp\left(-\frac{\pi}{2\sqrt{2}} \frac{1}{\sqrt{b-b_c}}\right). \quad (11)$$

즉 무한차 상전이 현상이 일어난다. (ii) 유한크기의 송이들의 평균크기 $\langle s \rangle$ 는 임계점에서 발산하지 않고 불연속적인 점프가 일어나게 된다. (iii) 임계점에서 송이 크기 분포함수는 $p(s) \sim 1/s^2 \ln^2 s$ 로 감소한다. 로그 보정이 있는 것이 보통의 경우와 다르다. (iv) 더욱 흥미로운 사실은 $b < b_c$ 에 대하여서도 $p(s) \sim s^{-\tau}$ 와 같은 멱함수 분포가 성립되고, τ 는 b 가 b_c 에서 0 으로 감소하면서 $2 \rightarrow \infty$ 로 증가한다. 이와 같은 특징들은 XY 모형의 보편성 군의 특징과 같다.

V. 좁은세상 네트워크

왓츠와 스트로가츠(WS)는 전력망, 선충(C. elegans)의 신경망, 영화배우 연결망에서 앞서 정의한 두 노드간의 평균거리와 결집계수를 측정하였다 [6]. 그 결과는 Table 1과 같다.

Table 1에서 볼 수 있는 바와 같이 결집계수는 실재하는 네트워크에서 측정된 값과 무작위 네트워크에서 측

정한 값이 큰 차이가 있음을 알았다. 이를 설명하기 위해서, WS는 다음과 같은 모형을 고안하였다. 원주 위에 N 개의 노드가 배열되어 있는 경우를 생각하자. 각 노드들은 최인접 이웃과 그 다음 최인접 이웃의 노드들과 연결되어 있다고 하자. 이 상태에서 각 연결선을 p 라는 확률로 선택하여 그 연결선의 양끝의 두 개의 노드 중 하나를 임의로 선택하여 고정하고, 다른 하나는 다른 노드로 재배열된다. 여기서 재배열되는 노드는 전체 네트워크에서 제한 없이 선택되게 된다. WS는 p 가 작은 경우 네트워크는 초기의 규칙적인 네트워크와 같이 큰 결집 계수를 가지고 있지만 두 노드간의 평균 거리는 무작위 네트워크처럼 매우 짧아진다는 것을 보여, 위의 실제 네트워크에서 얻은 결과를 만족시킨다고 주장하였다. 이러한 인위적인 네트워크 위에서 여러 가지 재미있는 물리현상을 연구한 논문들이 많이 발표되었다. WS의 논문의 제목에서 비롯하여 WS 모형을 좁은세상 네트워크라고 좁은 의미로 부르지만, 넓은 의미에서는 무작위 네트워크, 후에 소개할 척도없는 네트워크까지를 포함하여 두 노드간의 평균거리가 N 의 멱함수보다 작게 축척되는 경우를 일컫는다.

VI. 척도없는-네트워크의 예

1. 월드-와이드 웹

월드-와이드 웹의 네트워크 구조는 1999년 알버트와 그녀의 공동연구자 [7] 및 후버만과 그의 공동연구자 [8]에 의해서 밝혀졌는데, 그들이 조사한 네트워크의 노드 수는 약 20 ~ 30만개 정도였다. 그 이후 매우 큰 사이즈의 웹-페이지 네트워크에 대한 조사가 이루어 졌다. 쿠마르와 그의 공동연구자 [36-39] 들은 약 4 천만개, 브로더와 그의 공동연구자 [40] 는 약 2 억개의 웹-페이지로 구성된 네트워크를 조사하였다. 노드 수가 많은 데이터를 분석함으로써 식 (1)의 멱함수가 넓은 k 의 영역에서 성립함을 보일 수가 있었다. 연구 결과 웹-페이지에서 나가는-연결선수 지수(out-degree exponent)는 계에 따라 달라지는데 반하여 들어오는-연결선수 지수(in-degree exponent)는 $\gamma_{in} \approx 2.1$ 로 네트워크에 구별 없이 거의 일정하다는 것을 알게 되었다. 현재 지구 상에는 약 8억개의 웹페이지가 존재한다고 추정하고 있는데, 현재까지의 연구 결과를 토대로 추정해보면 8억개의 웹 페이지들 중에서 임의의 두 웹 페이지 간의 평균 거리는 19라고 추정하였다 [7]. 즉 19번 클릭만 하면 모든 지구상의 웹 페이지에 도달할 수 있다는 의미가 된다. 그러나 이

와 같은 연구결과는 엄밀한 결과는 아니다. 월드-와이드 웹에서는 연결선이 방향을 가지고 있기 때문에 도달할 수 없는 노드들이 존재하게 된다. 따라서 두 노드 간의 평균거리가 얼마나 되는가를 구하는 것은 엄밀하게 잘 정의된 양은 아니다. 평균 거리 19의 결과는 월드-와이드 웹의 각 연결선에 존재하는 방향을 무시하고 연결선이 있으면 양 방향으로 모두 갈 수 있다는 가정 하에 구한 두 노드간의 평균거리, 즉 지름을 잰 결과이다. 참고로 월드-와이드 웹의 유사성계수는 $r_k \approx -0.065$ 이다.

2. 인터넷

1999년 팔루소스 3형제는 인터넷에 대한 네트워크 구조를 연구하였다 [41]. 인터넷에서 노드를 어떻게 설정할 것인가는 정의하기에 달려 있다. 인터넷의 구조를 이해하는 단위로 종종 사용되는 노드는 도메인 또는 autonomous system(AS)이라고 부르는 것인데 예를 들어 학교 외부와 통신하는 서울대학교의 대표 라우터 정도의 레벨이 하나의 기본적인 단위가 되어 한 노드로 표현된다고 생각하면 된다. 월드-와이드 웹과는 달리 인터넷에서는 연결선이 양방향으로 통신할 수 있다고 생각하고 있다. 물론 현실적으로는 방향에 따라 정보처리 속도(bandwidth)가 달라 방향성이 있다고 생각할 수 있지만 문제를 간단히 만들기 위해 양방향의 처리 속도가 같다고 생각하자. 전세계 인터넷 구조는 미국의 Oregon 대학과 국립 응용 네트워크 연구 실험실(National Laboratory for Applied Network Research)에서 매일 파악하고 있다. 이 정보는 <http://moat.nlanr.net>에서 다운로드하여 볼 수 있다. 이 인터넷 AS 구조에 대한 정보는 1997년 11월부터 매일 저장되고 있다. 팔루소스 형제가 조사한 연구결과에 따르면 AS를 노드로 갖는 네트워크에서의 연결선수의 분포는 멱함수를 따르고 그의 지수는 $\gamma \approx 2.15 \sim 2.2$ 라는 것을 알게 되었다. 좀더 낮은 레벨인 라우터가 노드가 되는 네트워크에서는 연결선수의 지수가 $\gamma \approx 2.48$ 이었다.

인터넷 AS의 네트워크에 대한 정보는 매일 수집되고 있으므로 네트워크가 어떻게 진화해 가는가를 연구하는 데 좋은 대상이 되고 있다. 최근 고와 그의 공동연구자가 수행한 연구 [42]에 따르면 인터넷 AS 계의 노드수 N 은 시간(월을 단위시간으로 하는)에 대하여 $N(t) = N(0) \exp(a_1 t)$ 로 증가한다. 여기서 $a_1 \approx 0.029(1)$ 이다. 또 연결선수는 $L(t) = L(0) \exp(a_2 t)$ 로 증가하는데, $a_2 \approx 0.034(2)$ 이다. 따라서 연결선수가 노드수에 비하여 빨리 성장하는 가속성장-네트워크이다. 인터넷 구

조를 조사해 보면 한번 연결된 노드들 간의 연결선은 시간이 지나면서 고정되어 있지 않고, 연결선이 끊어지고 다른 노드로 연결되는 재배열 (rewiring) 현상이 심각히 일어나고 있다. 이러한 재배열 현상은 성장과정에서의 잡음 (noisy) 효과로 이해할 수 있고, 따라서 확률 과정 (stochastic process)을 이용한 분석 방법을 이용하여 네트워크 구조의 시간적 진화현상을 이해할 수 있다.

인터넷의 특징은 결집계수가 $C \approx 0.18 \sim 0.3$ 정도로 같은 노드 수, 같은 연결선수로 만든 무작위 네트워크의 결집계수 $C \approx 0.001$ 정도 보다 매우 높다. 이렇게 결집계수가 큰 이유는 집단구조 (community structure)를 이루고 있기 때문이다. 인터넷이 광케이블로 연결되어 있기 때문에 지역적인 중계소의 통제하에 있게 되어 집단구조를 만들고 있다. 따라서, 계층 결집계수 $C(k)$ 는 $\sim k^{-1}$ 와 같이 행동한다 [43]. 연결선수 사이의 상관관계는 $\langle k_{nn} \rangle \sim k^{-0.5}$ 로 연결선수가 많은 노드는 연결선수가 적은 노드와 연결된다 [30]. 유사성 계수도 $r_k \approx -0.189$ 로 측정되어 위의 성질을 뒷받침한다. 이러한 인터넷의 구조적 특성을 이해하는 방법으로 고와 그의 공동연구자는 적응 모형이라는 것을 제시하였다. 이 모형은 제 VII5 절에서 소개하기로 하겠다.

3. 영화배우 네트워크

영화배우 네트워크는 사회 네트워크의 한 예로 많은 연구의 대상이 되고 있다. 최근의 데이터들은 약 40 만 명의 영화 배우들로 구성되어 있다 [9,44,45]. 영화배우 네트워크는 영화배우와 그들이 출연한 영화가 함께 조합되어 이루는 네트워크이다. 이 네트워크는 두 종류 (A 형과 B 형)의 노드들로 구성되어 있다. A 형 노드들은 영화배우들이고 B 형 노드들은 영화에 해당된다. 같은 종류의 노드들은 서로 연결하지 않고, 다른 종류의 노드들끼리만 연결하게 된다. 이런 네트워크를 이중분할 (bipartite) 네트워크라고 한다. 이러한 두 종류의 노드를 가진 네트워크에서 한 종류만의 노드를 가진 네트워크로 전환하는 방법으로 흔히 쓰이는 것은 같은 영화에 출연한 영화배우들은 서로 모두 연결되어 있다고 생각하는 것이다. (즉 어느 한 B 형 노드에 연결된 A 형 노드를 모두 연결되어 있다고 하는 것이다.) [10] 이렇게 구성된 것이 영화배우 네트워크이다. 이 영화배우 네트워크의 연결선수의 분포함수는 정확히 멱함수를 따르고 있지는 않다. 그 이유는 한 영화배우가 일생동안 무수히 많은 영화에 출연할 수 없기 때문이다. 이러한 성질은 사회네트워크의 일반적 특징이라고 할 수 있다. 그렇지만 지수함수적

으로 감소하는 무작위 네트워크와 비교하면 연결선수가 상대적으로 많은 노드가 존재하는 뚱뚱한-꼬리 (fat-tail) 현상을 보이고 있다고 할 수 있다. 영화배우 네트워크는 사회연결망을 연구하는데 좋은 예로 종종 사용된다. 왜냐하면 우선 노드 수가 많고, 데이터들이 정확하기 때문이다. 또 유사성 계수값 $r_k \approx 0.208$ 도 양의 값을 보인다.

4. 논문 인용 네트워크

이 네트워크에서는 노드가 논문이 된다. A 라는 논문이 B 라는 논문을 인용하게 되면 A 라는 노드에서 B 라는 노드로 화살표가 가는 연결망이 구축된다. 즉 방향이 있는 네트워크가 된다. 하루에도 많은 논문이 쏟아져 나오고 있으므로, 이 네트워크는 노드 수가 정지되어 있지 않고 성장하는 네트워크가 된다. 새로 태어나는 논문은 기존의 논문을 인용하게 된다. 따라서 연결선수가 많은 노드는 오래된 논문일 것임을 간단히 예상할 수 있다. 이 논문 인용 네트워크는 월드-와이드 웹이 척도없는-네트워크라는 것이 소개되기 전에 레드너가 1998년 논문 인용 횟수를 조사하여 밝혀진 것이다 [46]. 그는 1981년부터 1997년 6월까지 ISI 논문인용 색인을 조사하였고, 1975년부터 1994년까지의 Physical Review D 의 논문의 인용 상황을 조사하였다. 어떤 논문이 k_i 번 인용될 빈도 $P_{in}(k_i)$ 를 조사한 결과 $P_{in}(k) \sim k^{-3}$ 의 곡선과 걸맞았고, 식 (2)와 같은 Pareto 식에 맞추어 보면 ISI 에 대해서는 $\gamma \approx 2.9$, Physical Review D 에 대해서는 $\gamma \approx 2.6$ 에 잘 맞았다 [47]. 그러나 k 가 매우 큰 영역에서의 분포함수가 급격히 감소하는 성질을 보여주는 경향이 있어 이에 대한 논란이 있다. 또한 인용하는 논문의 참고문헌 수의 분포 함수는 네트워크의 나가는-연결선수의 분포함수를 의미하게 되고 이에 대한 조사 결과는 멱함수가 아닌 꼴이 나왔다 [48].

5. 신진대사 반응 네트워크

일반적으로 일련의 화학반응을 그래프로 그려보는 것은 화학반응의 흐름을 알기에 유용하다 [49]. 신진대사 반응 네트워크 (metabolic reaction network)는 생명체가 살아가는 데 필요한 에너지를 공급하기 위해 세포속에서 작용하는 일련의 화학반응인 신진대사 반응을 네트워크로 표시한 것이다. 예를 들어 Fig. (1a)와 같은 화학반응이 있다고 하였을 때, 이에 대한 네트워크를 Fig. (1b)와 같이 그릴 수 있다. 이러한 네트워크는 이중분

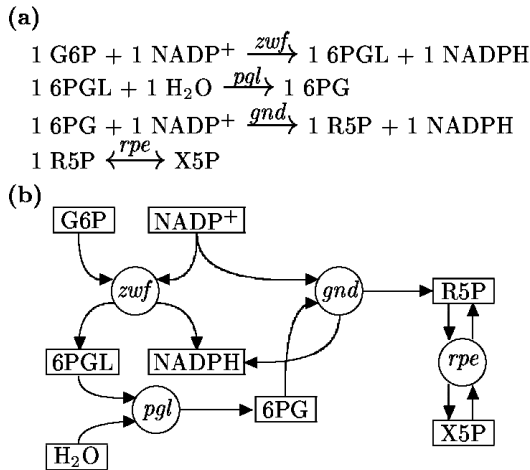


Fig. 1. A few stoichiometric equations for metabolic reactions (a) and their corresponding network representation (b). Rectangle represents metabolite and circle does enzyme.

할(bipartite) 그래프이다. 한 종류의 노드는 화학반응에 참여하는 화학물(metabolite)이고, 다른 한 종류의 노드는 화학반응에 관련된 촉매이다. 이러한 신진대사 네트워크의 연결선수의 분포 함수가 멱함수 꼴을 따름을 보인 것은 정과 그의 공동연구자 [50], 그리고 와그너와 펠 [51]이었다. 정 등은 43 가지의 원핵생물, 박테리아, 진핵생물에서 일어나는 신진대사 네트워크의 연결선 분포에 대한 연구를 수행하여 척도없는 성질이 있음을 밝혔다. 또한 43 개의 종류들은 진화 정도의 차이에 의하여 노드의 수에 많은 차이가 있는데도 두 점간의 평균거리가 일정하다는 흥미로운 결과를 얻었다. 그러나 최근의 연구 결과에 의하면, ATP, NADP 와 같은 신진대사 반응에서 공통적으로 참여하는 필수 화합물들을 제거하여 네트워크를 구성하면, 두 점간의 평균거리는 노드의 수에 의존한다는 사실이 밝혀졌다. [52].

6. 단백질 상호작용 네트워크

단백질 상호작용 네트워크에 대한 연구도 활발히 일어나고 있다. 생리현상이 최종적으로 단백질에 의하여 표출되므로 질병을 일으키는 생물체에 대한 단백질 상호작용 네트워크에 대하여 관심이 크다. 현재 생물학계에서는 단백질 각각의 기능을 연구하는 것보다는 단백질들이 연결되어 일으키는 총체적인 결과를 이해하여야 한다는 측면이 강조되고 있다. 따라서 단백질 상호작용 네트워크에 대한 이해가 필요시 되고 있다. 특히 게놈 프로젝트 [53]로 인하여 유전자의 염기 서열이 밝혀지면서, 이 염기 서열에서 유추되는 3 차원적인 단백질 구조

를 추측하게 되었다 [54]. 단백질의 3 차원 구조는 단백질 상호작용을 일으키는 주요한 역할을 하게 된다. 최근 단백질 간의 상호작용 여부를 판단하는 이중의 하이브리드 방법이 개발되어 많은 단백질 간의 상호작용 여부를 판단할 수 있다 [55,56]. 가장 간단한 원소인 수소, 헬륨이 물리학에서 기본 모형이 되는 것과 같이 E. Coli 와 효모(yeast)에 대한 단백질 상호작용 네트워크에 대한 연구가 최근 생물학계와 물리학계에서 수행되었다. 이에 대하여 살펴보기로 하겠다. 이스트에 대한 이중 하이브리드 방법을 이용한 단백질 상호작용에 대한 연구는 1999 년 Uetz 등에 의해 수행되었다 [57]. 초기에는 6000 여개의 이스트 단백질 중에 약 1000 여개의 단백질만의 상호작용에 대하여 연구되었다. 이후 Ito 등에 의하여 좀 더 많은 단백질들에 대한 상호작용 연구가 수행되었다 [58,59]. 그 이후 최근에는 DIP 와 MIPS 등의 데이터베이스에는 5000 여개의 단백질간의 상호작용에 대한 데이터가 축적되어 있다 [60,61]. 단백질 상호작용 네트워크에서는 단백질이 노드가 되고 두 단백질이 서로 상호작용을 하는 경우 연결선이 있다고 가정한다.

이렇게 구성된 이스트 단백질 상호작용 네트워크의 특징은 다음과 같다. 첫째, 이스트 단백질 상호작용 네트워크의 연결선수 분포 함수는 일반적인 멱함수 꼴, 즉 Pareto 함수, 식 (2)를 따른다. Uetz 등의 데이터를 기반으로 하는 단백질 상호작용 네트워크에서는 데이터가 작기 때문에 지수함수의 자름(cutoff)이 있는 멱함수 꼴이라는 제안 [62]도 있었으나, 최근에는 Pareto 식에 따른다는 것이 일반적인 견해이다 [28]. 둘째, 이스트 단백질 상호작용 네트워크에서 연결선수가 많은 단백질은 치사(lethal) 단백질일 확률이 높다 [62]. 연구에 따르면 약 70 % 가까이 된다는 것이 알려져 있다. 여기서 치사 단백질이란 그 단백질이 제거 되었을 때 이스트가 생존할 가능성이 없는 단백질을 의미한다. 셋째, 이스트 단백질은 모듈 구조를 가지고 있다. 같은 기능을 가지고 있는 단백질들은 네트워크에서 지역적으로 모여 있는 모듈을 형성하고 있다. 따라서, 결집계수가 비교적 높은 편이다. $C \approx 0.13$. 모듈 형성은 결집계수의 연결선수의 의존성 $C(k)$ 를 통해 알 수 있다. $C(k)$ 는 k 가 작은 경우에는 거의 상수이고 k 가 큰 경우에는 $C(k) \sim k^{-1}$ 로 감소한다 [29]. 넷째, 연결선수가 적은 노드들은 연결선수가 많은 노드들과 연결되어 있다. k 의 연결선수를 가지고 있는 노드의 이웃 노드들의 평균 연결선수 $\langle k_{nn} \rangle(k)$ 는 k 가 증가할수록 감소하는 감소함수 성질을 보인다 [64]. 즉 비유사성의 성질 $r_k \approx -0.16 < 0$ 이 있다.

VII. 척도없는-네트워크에 대한 모형

척도없는-네트워크에 대하여 지금까지 많은 모형이 소개되었다. 그 중 중요한 몇 가지 모형을 소개하기로 하자.

1. 바라바시-알버트(BA) 모형

척도없는-네트워크가 어떻게 생겨나게 되었는지에 대하여 바라바시-알버트(BA)는 다음과 같은 모형을 소개하였다 [45]. 이 모형은 두 가지 중요한 요소를 포함하고 있다. 첫째, 이 모형은 성장하는 네트워크이다. 즉 네트워크의 노드 수는 시간이 지남에 따라 증가한다. BA는 편의상 단위시간당 한 개의 노드가 새로이 생겨난다고 가정하여 시간 t 가 지난 후 네트워크에 존재하는 총 노드 수는 $N = t$ 이다. 둘째는 새로이 생겨난 노드는 m 개의 연결선을 기존의 노드에게 연결시킨다. 이때 연결되는 노드는 무작위로 선택되는 것이 아니라 그 시각에 존재하는 각 노드의 연결선수에 비례하여 선택된다. 이를 선호적 연결(preferential attachment)이라고 한다. 즉 t 라는 시각에 태어난 노드가 m 개의 연결선을 기존의 노드 i 에 연결할 확률은

$$\Pi_i = k_i(t) / \sum_j k_j(t) \quad (12)$$

로 주어진다. 여기서 $k_i(t)$ 는 시각 t 에 i 노드의 연결선수이다. 이러한 연결선수의 변화를 식으로 기술하면

$$\frac{\partial k_i}{\partial t} = m\Pi_i = m \frac{k_i(t)}{\sum_j k_j(t)} \quad (13)$$

이다. $\sum_j k_j \approx 2mt$ 라는 성질을 이용하면,

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}. \quad (14)$$

이 방정식의 해는 $k_i(t) = m(t/t_i)^{1/2}$ 이다. 여기서 t_i 는 노드 i 가 태어난 시각이다. 노드의 연결선수 $k_i(t)$ 가 k 보다 작을 확률은

$$P(k_i(t) < k) = P_d(t_i > m^2t/k^2). \quad (15)$$

그런데 단위 시간 동안 한 개의 노드가 네트워크에 첨가되므로 $P(t_i) = 1/(m_0 + t)$ 이다. 여기서 m_0 은 초기에 존재하는 노드 수이다. 따라서, 식 (15)의 우변은

$$P(t_i > m^2t/k^2) = 1 - \frac{m^2t}{k^2(t + m_0)}. \quad (16)$$

그러므로 연결선수 분포함수 $P_d(k)$ 는

$$P_d(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^2t}{m_0 + t} \frac{1}{k^3}. \quad (17)$$

따라서, $t \rightarrow \infty$ 의 경우에는, $P_d(k) \sim 2m^2k^{-3}$ 이다. 이 BA 모형은 다른 방법으로도 해석적인 해를 구할 수 있는데, 예를 들어 도로고체프와 그의 공동연구자 [65]들은 으뜸방정식(master equation) 방법을 사용하여

$$P_d(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \quad (18)$$

를 구했으며, 크래피프스키와 그의 공동연구자 [67]는 속도식(rate equation)을 사용하여 식 (18)을 얻었다.

BA 모형을 일반화하는 시도가 이루어졌다. 선호적 연결을 주는 확률 식 (12) 대신에 $\Pi_i = (k_i + ma) / \sum_i (k_i + ma)$ 로 주어지는 경우에는 연결선수 지수는 $\gamma = 3 + a$ ($a < 0$ 일 수 있음)로 주어진다 [65]. 이때 주의하여야 할 점은 $a < 0$ 일때, 연결선수 분포함수에서 k 가 큰 영역에서 혹(bump)과 같은 것이 생기게 된다는 것이다. 물론 이 혹은 N 이 유한한 성질 때문에 생기는 것이지만 시뮬레이션이 가능한 영역에서 심각한 영향을 발휘하기 때문에 주의하여야 한다 [66]. 또 $\Pi_i = k_i^y / \sum_i k_i^y$ ($y \neq 1$)의 경우도 연구되었는데 자세한 내용은 참고문헌 [67]을 참고하기 바란다.

2. 적합성 모형

적합성 모형은 BA 모형을 변형한 것으로 선호적 연결의 확률은

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}. \quad (19)$$

여기서 η_i 는 i 노드가 갖는 적합성(fitness)이다 [68,69]. 이 적합성은 노드가 태어날 때부터 부여받는 것으로 어떤 분포함수 $\rho(\eta)$ 를 따른다. 이 때 i 노드의 연결선수 k_i 의 시간에 대한 변화는 다음과 같은 동역학식에 의해서 기술된다.

$$\frac{\partial k_i}{\partial t} = m \frac{\eta_i k_i}{\sum_j \eta_j k_j}. \quad (20)$$

여기서 얻는 연결선수 분포함수는 적합성에 대한 분포함수 $\rho(\eta)$ 에 많이 의존한다. 만약 $\rho(\eta)$ 분포함수가 $[0,1]$ 사이에서 균일하게 분포되어 있는 경우에는 연결선수 분포는 $P(k) \sim k^{-2.26} / \log(k)$ 이다. 또, $\rho(\eta) = \exp(-\eta)$ 인 경우에는 펼쳐진 지수 함수를 따른다. 이 적합성 모형은

BA 모형과는 달리 $\langle k_{nn} \rangle \sim k^{-\nu}$ ($\nu > 0$) 와 같이 행동하여, 즉 연결선수가 큰 노드는 연결선수가 작은 노드와 연결되는 성질이 있어서, 좀 더 실제의 네트워크에 가깝다는 것이 알려져 있다 [30].

이 모형에서 특별히 재미있는 것은 적합성 변수 η_i 를 $\epsilon_i = -k_B T \log \eta_i$ (k_B 볼츠만 상수, T 온도)의 형태로 치환하여 식 (20) 을 다시 쓰면

$$\frac{\partial k_i}{\partial t} = m \frac{e^{-\epsilon_i/k_B T} k_i}{Z_N} \quad (21)$$

로 쓸 수 있는데, 여기서 $Z_N = \sum_{j=1}^N e^{-\epsilon_j/k_B T} k_j$ 로 평형 통계학에서 분배 함수와 같은 꼴이 된다. 좀더 수학적인 조작을 통하면 적합성 ϵ 를 가지고 있는 노드들에 연결된 연결선의 수는

$$n(\epsilon) = \frac{1}{e^{(\epsilon-\mu)/k_B T} - 1} \quad (22)$$

로 주어진다. 여기서 μ 는 휘산도인데,

$$e^{-\mu/k_B T} = \lim_{N \rightarrow \infty} \frac{\langle Z_N \rangle}{m t} \quad (23)$$

로 주어진다. 식 (22) 은 잘 알려진 보즈 통계와 같은 것으로 적합성(η)이 매우 큰 노드, 즉 에너지 ϵ 이 작은 노드에 많은 연결선이 연결되는 것은 에너지가 작은 상태(state)에 많은 보즈 입자가 쌓이게 되는 것과 같은 이유라고 할 수 있다.

3. 후버만-아다미치 모형

후버만과 그의 공동연구자 아다미치(HA)는 BA 모형과는 다른 모형을 소개하였다 [8]. HA 모형에 따르면 계에 존재하는 노드들의 개수 $N(t)$ 는 시간에 대하여 지수적으로 증가한다. 즉 $N(t) = N(0) \exp(a_{HA} t)$ 이다. 시각 t 에서 노드 i 의 연결선수 $k_i(t)$ 는 다음 시각 $t+1$ 에서 다음과 같이 변화한다.

$$k_i(t+1) = k_i(t)(1 + \zeta_i(t+1)), \quad (24)$$

여기서 $\zeta_i(t)$ 는 시각 t 에서 증가율이다. 이 증가율은 시각에 따라 변화한다. 따라서 증가율을 두 부분으로 나눈다. 즉

$$\zeta_i(t) = g_{0,i} + \xi_i(t), \quad (25)$$

여기서 $g_{0,i}$ 는 노드 i 의 연결선수의 증가율을 의미하는데 시간에 의존하지 않고, $\xi_i(t)$ 는 잡음항으로 $\langle \xi_i(t) \rangle = 0$ 과 $\langle \xi_i(t) \xi_j(t') \rangle = \sigma_{0,i}^2 \delta_{t,t'} \delta_{i,j}$ 이다. 여기서 $\sigma_{0,i}^2$ 는 분산

이다. $\langle \dots \rangle$ 는 표본 평균을 의미한다. 편의상 증가율의 로그함수를 $G_i(t) \equiv \ln(1 + \zeta_i(t))$ 로 표시한다. HA 는 $g_{0,i} = g_0$ 와 $\sigma_{0,i} = \sigma_0$ 를 가정하였다. 편의상 노드 지표 i 에 의존하지 않는다고 생각하였다. 그러면 $t = t_0$ 에서 $k_i = k_0$ 이면서 시간 $t = t_0 + \tau$ 에 $k_i = k$ 일 조건 확률 $P_D(k, \tau | k_0)$ 은 다음과 같이 주어진다.

$$P_D(k, \tau | k_0) = \frac{1}{k \sqrt{2\pi\sigma_{\text{eff}}^2 \tau}} \exp \left\{ -\frac{(\ln(k/k_0) - g_{\text{eff}}\tau)^2}{2\sigma_{\text{eff}}^2 \tau} \right\}, \quad (26)$$

여기서 $g_{\text{eff}} \equiv \langle G_i(t) \rangle$ 와 $\sigma_{\text{eff}}^2 \equiv \langle (G_i(t) - \langle G_i(t) \rangle)^2 \rangle$. 또 g_{eff} 와 σ_{eff}^2 는 다음과 같은 관계식을 갖는다. $g_{\text{eff}} \approx g_0 - \sigma_0^2/2$ 과 $\sigma_{\text{eff}}^2 \approx \sigma_0^2$ [70].

나이 τ 를 가지고 있는 노드들의 밀도는 $\rho(\tau) \sim \exp(-a_{HA}\tau)$ 이므로, k 라는 연결선수를 가지고 있는 노드들의 수는 모든 나이들에 대한 합을 구하여야 한다.

$$P_d(k) \sim \int d\tau \rho(\tau) P_D(k, \tau | k_0) \sim k^{-\gamma}, \quad (27)$$

여기서,

$$\gamma = 1 - \frac{g_{\text{eff}}}{\sigma_{\text{eff}}^2} + \frac{\sqrt{g_{\text{eff}}^2 + 2a_{HA}\sigma_{\text{eff}}^2}}{\sigma_{\text{eff}}^2}. \quad (28)$$

그러므로 연결선수 지수 γ 는 g_{eff} 와 σ_{eff} 에 의해서 결정된다. 최근 이 모형을 기본으로 하여 인터넷의 진화현상을 설명하려는 연구가 수행되었다 [42].

4. 스택틱 모형

스택틱 모형은 ER의 무작위 네트워크 모형을 좀더 일반화한 것으로 성장모형이 아니다. 즉 ER 모형처럼 처음부터 N 개의 노드가 존재하고 노드들은 $i = 1 \dots N$ 으로 표시된다. 각 노드에 $w_i = i^{-\alpha}$ 의 가중치를 부여한다. 이 가중치에 따라 연결선이 연결된다. 즉 임의의 두 개의 노드 (i, j) 를 $p_{i,j} = i^{-\alpha} j^{-\alpha} / \sum_{\ell, m} \ell^{-\alpha} m^{-\alpha}$ 의 확률로 선택한다. 만약 그 두 개의 노드들 사이에 연결선이 존재하면, 다시 다른 두 개의 노드를 각 노드의 가중치의 곱에 비례하는 확률로 선택하여 연결선을 잇는다. 이런 방법으로 계속 연결선을 붙여가서 L 개의 연결선이 될 때까지 연결한다. 만약 α 가 영이면 이 모델은 가중치가 없는 경우가 되며, 이것이 바로 ER 모형이 된다. 이 모형은 고와 그 공동연구자가 [71] 소개하였으며, 처음 소개되었을 때에는 BA 모형과 유사성을 취하기 위해 $L = mN$ (m 은 자연수)의 경우를 생각하였다. 이 경우 N 개의 노드들은 모두 연결되며, 연결선수의 분포함

수는 멱함수를 따르게 된다. 후에 이태리 연구팀에서도 유사한 모형이 연구되었다 [72].

연결선수 분포함수를 생각해 보자. 노드 i 가 연결선수 k 를 가질 확률을 $p_{d,i}(k)$ 라 표현하면, $p_{d,i}(k) \simeq \langle k_i \rangle^k \exp(-\langle k_i \rangle)/k!$ 로 쓸 수 있는데, 이것은 i 라는 노드가 선택될 확률은 $i^{-\alpha}$ 에 비례하지만 선택 과정이 확률적이므로 푸아송 분포를 따르게 된 것이다. 그런데, $\langle k_i \rangle = K(L)i^{-\alpha}$ 이다. 여기서 $K(L) = \langle k_1 \rangle = 2L/\zeta_N[\alpha]$ 이며 $\zeta_N(x) = \sum_{j=1}^N j^{-x}$ 이다. 연결선수 분포 함수는

$$P_d(k) = \frac{1}{N} \sum_{i=1}^N p_{d,i}(k) \simeq ck^{-\gamma} \quad (29)$$

여기서 $c \simeq (\gamma - 1)[K(L)]^{\gamma-1}/N$ 이며, $\gamma = 1 + 1/\alpha$ 가 된다. 즉 변수 α 를 조절함으로써 연결선수 지수 γ 가 다른 척도없는-네트워크를 만들 수 있다. 이 스테틱 모형의 특징은 BA 모형이 $2 < \gamma < 3$ 인 경우 k 가 큰 영역에서 연결선수 분포 함수가 멱함수 꼴에서 벗어나는 성질이 있는 반면, 이 스테틱 모형은 k 가 큰 영역에서도 연결선수 분포 함수가 멱함수 꼴에 잘 맞는 장점을 가지고 있다. 따라서, 척도없는-네트워크 위에서 다양한 문제를 시뮬레이션하기 위해 우선 네트워크를 만들어야 하는 경우 지금까지 나온 여러 모형 중에서 이 모델이 가장 유용하다. 그러나 연결선수 간의 상관관계가 없는 경우라서 실재하는 네트워크의 성질을 조사하기 위해 사용하는 데는 적합하지 않다.

5. 적응 모형

앞에서 살펴본 BA 모형이나 스테틱 모형에서는 연결선이 노드에 일단 연결되면 그 연결은 변하지 않는다. 그러나 인터넷과 같은 실제 네트워크에서는 연결선은 어떤 노드에 연결되었다가도 시간이 지나가면 그 연결이 끊어지고 다른 노드로 연결되기도 한다. 이런 현상을 재배열(rewiring) 현상이라고 한다. 재배열 현상에 대한 모형은 알버트-바라바시가 소개하였다 [73]. 그 모형에서는 네트워크가 다음과 같은 규칙에 따라 만들어진다. (i) p 의 확률로 기존의 노드 중 한 노드를 선택한 다음 그 노드에서 m 개의 연결선을 m 개의 서로 다른 노드를 선호적으로 선택하여 연결한다. (ii) q 의 확률로 어떤 연결선을 선택하여 양쪽에 있는 노드 중의 하나는 고정하고 다른 쪽에 있는 노드에서 연결선을 떼어낸 후 다른 노드로 재배열한다. 이런 재배열 작업을 m 번 반복한다. (iii) $1 - p - q$ 의 확률로 새로운 노드를 네트워크에 입력시키고 이 노드로부터 m 개의 연결선을 기존의 m 개의 노

드에 선호적 연결 방법에 따라 연결시킨다. 이러한 방법에 따라 만들어진 네트워크는 Pareto 형의 일반적인 멱함수 꼴을 따르게 된다. 이때 연결선수의 지수는 변수 p, q 에 의존하게 된다. 재배열 모형은 변수에 따라 연결선수 지수를 조정할 수 있다는 장점이 있으나 근본적으로 네트워크의 구조적 특징, 즉 연결선수 간의 상관관계라든지, 결집계수 등을 크게 바꾸지 못한다.

최근 적응 모형(adaptation model)이라는 것이 소개되었다 [42]. 이 모형의 기본 아이디어는 재배열하는 과정에서 즉 (ii)의 연결선이 A 라는 노드에서 떨어져 B 라는 노드로 연결되었을 때, B 노드의 연결선수 k_b 가 A 노드의 연결선수 k_a 보다 크다는 조건을 만족시키도록 하는 것이다. 이 모형에는 노드가 재배열하는 것은 효율성을 높이기 위해서라는 아이디어가 담겨있다. 그렇지 않으면 노드가 현실적으로 재배열할 필요가 없기 때문이다. 다른 관점에서 생각해 보면 노드가 시간이 지남에 따라 환경이 바뀌게 되므로 더 좋은 환경으로 적응한다고 볼 수도 있다. 적응 모형에서는 이런 요소를 고려한 것이다. 또, B 노드는 A 노드의 이웃에서 고른다는 조건을 부여하게 되면, 이 조건은 인터넷 등에서도 같이 지역적으로 새로운 환경이 만들어지는 경우 그 좋아진 환경으로 재배열이 이루어 지는 것을 뜻하게 된다.

이러한 적응요소가 첨가된 재배열 과정이 포함된 모형은 인터넷과 같은 실제 존재하는 네트워크의 구조적인 특징, 즉 결집계수 값이 크든지, 작은 연결선수를 가진 노드와 많은 연결선수를 가진 노드가 연결된다는 등의 특징을 만족시킨다. 따라서, 적응현상이 복잡계 네트워크에서 상당히 중요하다는 것을 알 수 있다.

6. 단백질 상호작용 네트워크 모형

우선 솔레 등이 소개한 모형을 살펴보자 [74]. 그들은 단백질 상호작용 네트워크를 구성하는 기본적인 3 가지 메커니즘을 다음과 같이 지적하였다. (i) 복제(duplication): 임의의 노드(단백질)를 선택하여 복제시킨다. 노드가 복제됨에 따라 노드뿐만 아니라 그 노드에 연결된 모든 연결선이 복제된다. 즉 복제 노드는 본디 노드가 연결된 모든 이웃 노드들과 연결된다. 이 복제는 단백질의 진화는 유전자 복제과정에서 비롯된다는 생각에서 연유한다. (ii) 다변화(divergence): (i)의 과정을 통해서 생긴 복제 노드와 본디의 노드가 병존하게 되는데 이는 똑같은 두 개의 단백질이 존재한다고 할 수 있다. 이러한 중복성을 피하기 위해 복제 노드에 연결된 각각의 연결선을 δ_{pin} 라는 확률로 제거한다. (iii)

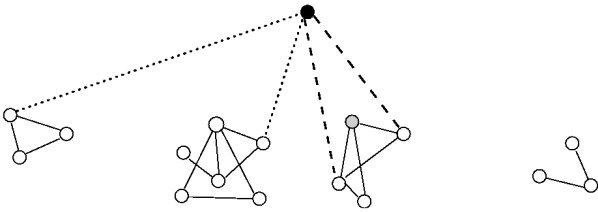


Fig. 2. Evolution steps of the protein interaction network model. (i) duplication: The new node (heavy dot) duplicates the 3 links between the target node (shaded) and its neighbors (dashed), (ii) divergence: but among them, one of the links is removed with probability δ_{pin} . (iii) mutation: The new node also attaches to any other network nodes (dotted) with probability α_{pin} .

돌연변이(mutation): 유전자 복제 과정에서 돌연변이가 일어나기도 한다. 이를 반영하기 위하여 복제 노드에서 본디의 노드의 이웃 노드 외의 노드 중 $\alpha_{pin}N$ 개를 선택하여 연결한다. 위의 세 가지 단계는 Fig. (2)에서 확인할 수 있다.

이와 같은 복제-다변화-돌연변이 3개의 요소가 단백질 상호작용 네트워크를 구성하는 기본 메커니즘이라는 것을 솔레 등은 제의하였다. 이 모형은 연결선수 분포가 일반적인 멱함수 꼴 (Pareto 형)을 따르는 단백질 상호작용 네트워크의 특징을 재현할 수 있지만, 결집계수, 연결선수 간의 상관관계 등의 구조적 특징을 재현하지는 못한다. 이에 대한 해석적인 이해는 다음 논문을 참고하기 바란다 [75].

다음, 바즈케즈 등이 제안한 모형을 소개하자 [76]. (i) 복제: 임의의 노드 A 를 선택하고 이의 복제 노드를 A' 라 하자. A 의 연결선도 복제한다. 추가적으로 본디의 노드 A 와 복제 노드 A' 사이는 p_{pin} 라는 확률을 가지고 연결선을 연결한다. (ii) 다변화: 노드 A 와 A' 의 공통 이웃인 B 로 부터 A 또는 A' 에 연결된 연결선을 q_{pin} 라는 확률로 제거한다. 이 모형에서는 돌연변이 과정이 생략되어 있으며, 복제 노드와 본디의 노드 간의 대칭성이 존재한다. 이 모형은 복제 노드와 본디의 노드가 연결될 확률이 있기 때문에 자연적으로 A, A' 와 B 가 삼각형을 이룰 확률이 존재하고 따라서 결집계수가 높아진다. 또 $\langle k_{nn} \rangle$ 도 k 에 대하여 감소하도록 q_{pin} 를 잡으면 실제의 단백질 상호작용 네트워크의 성질을 잘 재현할 수 있지만 그렇게 하였을 때, 다변화되는 비율이 너무 높아져 모형에서의 큰 송이의 크기가 실제하는 큰 송이의 크기에 비하여 상대적으로 작아지게 되어 실제 네트워크에 어긋나게 된다.

이와 같은 모형 외에도, 베르그 등이 제의한 모형 [77]이 있으며, 쿠닌 [78]등이 소개한 모형도 있지만 각각을 실제 이스트 단백질의 상호작용 네트워크와 비교

해 보면 완전하지 못함을 알 수 있다. 최근 본 연구팀에서는 이러한 결점을 보완하는 새로운 모형을 고안하였다. 이 모형에서는 단백질 상호작용 네트워크에서 고려하여야 할 새로운 사항을 소개하였다. 단백질 상호작용은 근본적으로 단백질 도메인에 의한 것이며 [79] 또한 단백질들은 단백질 군 (protein family) [80]을 이루기 때문에 모형을 만들 때 이런 성질에 대한 고려가 필요하다는 것이었다. 단백질 군에 속해있는 단백질들은 거의 같은 구조를 가지고 있다는 것이 알려져 있다. 따라서, 돌연변이가 이루어진다고 하더라도 크게 단백질 군을 벗어나지는 않는다는 것이 생물학적으로 알려진 사실이다. 이러한 단백질 군의 존재는 앞서 설명한 네트워크의 모듈 구조의 존재를 의미하는 것과 같은 아이디어이다. 솔레 모형을 약간 변형하여 단백질 상호작용은 단백질 도메인에 의해 일어나며, 돌연변이 과정 (iii)은 단백질 군 안에서만 일어난다고 하는 제한 조건을 사용하면 실제 단백질 상호작용 네트워크의 구조적 특징을 잘 설명해 준다는 사실을 알게 되었다 [81].

7. 사회네트워크에 대한 모형

사회네트워크는 다른 복잡계 네트워크와는 다른 몇 가지 특징을 가지고 있는데, 이들은 다음과 같다. (i) 사회 네트워크에서는 총 노드 수가 시간이 지나면서 늘어나는 것이 아니라 정체되어 있다고, 즉 총 노드 수가 고정되어 있다고 생각하는 것이 좋다. 사람들이 태어나고 죽음에 따라 노드 수가 변하지만, 사람과 사람과의 관계 형성, 즉 새로운 사람과 만나 새로운 연결선이 형성되고, 시간이 지남에 따라 두 사람 관계는 소원해져 연결선이 끊어지는 시간의 크기 정도가 훨씬 짧아 노드 수가 고정되어 있다고 보는 것이 합리적이라고 할 수 있다. (ii) 연결선수의 분포가 멱함수를 따르지 않고, 최빈값이 존재한다. 즉 평균적으로 사람들이 만나는 수가 정해져 있다고 할 수 있다. 이렇게 멱함수를 따르지 않는 것은 선택적 연결이라는 요소가 존재하지 않는다는 것을 의미하기도 한다. (iii) 사람들 사이에서는 유유상종성, 또는 결집성이 매우 높다. 즉 끼리끼리 모으려는 경향이 뚜렷이 있다. 또 집단구조를 만들려는 경향이 뚜렷이 있다. 위의 특징이 모든 사회연결망에서 모두 나타나는 공통된 특징은 아니다. 예를 들어 영화배우 네트워크는 많은 신인배우들이 등장하기 때문에 성장하는 네트워크라고 할 수 있다. 위의 특징은 뉴우만이 제의한 것으로 좀 더 넓은 범위의 사회연결망에 적용될 수 있겠다.

이러한 특징을 설명하는 여러 모형이 있다. 진(Jin)등이 소개한 모형을 보면 [82] 사회연결망을 구성하는 중요한 요소로 다음 3 가지를 꼽았다. (i) 두 사람이 한 사람 이상의 사람을 공통적으로 알고 있으면 두 사람이 모임을 가질 확률이 높다. (ii) 자주 만나지 않는 두 사람 사이의 시간이 지날수록 관계가 멀어진다. (iii) 한 사람이 유지할 수 있는 친분관계에 있는 사람의 수는 제한이 있다. 이러한 요소를 가지는 네트워크를 컴퓨터 시뮬레이션을 통하여 재현하였더니 실제의 사회네트워크의 특징을 잘 설명해 주었다.

여섯단계 분리에 대한 모형이 최근 소개되었다 [83, 84]. 그들은 사회연결망을 사회그룹의 계층적 나무가지 구조를 만들어 거리를 정의했는데, 여기서 한 개인은 여러 그룹에 속해 있을 수 있기 때문에 사람들 사이의 거리는 계층구조에 의해 계산될 수 없는 더욱 좁은 세상을 만든다고 발표했다. 이와 관련하여 최근의 본 연구팀에서 발표한 논문에서는 각 개인은 여러 그룹에 속해 있을 수 있으므로 각 그룹에서의 적합성을 하나의 벡터의 성분으로 표시하는 모형을 소개하였는데, 이것은 앞서 소개한 스테틱모형을 q -성분을 가진 모형으로 일반화한 것에 해당된다 [85].

VIII. 척도없는-네트워크의 울트라 좁은 세상

무작위 네트워크와 WS 의 좁은 세상 네트워크에서의 네트워크의 총 노드의 수가 N 인 경우 두 노드간의 평균 거리는 $D \sim \ln N$ 이라는 것이 알려져 있다. 최근 척도없는-네트워크에서 두 노드 사이의 평균거리는 이보다 짧은 울트라 좁은 세상의 성질이 있다는 것이 밝혀졌다 [86, 87]. 결과만 살펴보면 연결선수 지수 γ 가 $2 < \gamma < 3$ 인 경우 $D \sim \ln \ln N$ 이고 $\gamma = 3$ 인 경우에는 $D = \ln N / \ln \ln N$ 이며 $\gamma > 3$ 인 경우에는 $\ln N$ 이다.

IX. 연결선수 간의 상관관계

연결선수의 분포 뿐만이 아니라 연결선수 간의 상관관계도 네트워크의 구조를 이해하는데 중요하다 [88]. 어느 한 연결선을 무작위로 선택하여 양쪽에 있는 노드를 i, j 라 하자. 두 노드의 연결선수 간의 상관함수 $P(k, k')$ 는 다음과 같이 기술된다.

$$P(k, k') = \frac{1}{N^2} \left\langle \sum_{i,j=1}^N \delta(k_i - k) \delta(k_j - k') \right\rangle. \quad (30)$$

만약 두 노드간에 상관관계가 없으면

$$P(k, k') = \frac{kP(k)k'P(k')}{(\sum_k kP(k))^2} \quad (31)$$

이다. 따라서, 만약 상관관계가 없는 척도없는-네트워크인 경우에는 $P(k, k') \sim k^{1-\gamma}k'^{1-\gamma}$ 가 된다. 스테틱 모형에서의 노드간의 상관함수가 이렇게 주어진다.

그러나 바라바시-알버트 모형이나 논문인용네트워크와 같이 성장하는 네트워크에서는 연결선수 상관함수는 $k \gg k' \gg 1$ 인 경우 $P(k, k') \sim k^{-(\gamma-1)}k'^{-2}$ 임을 크래피프스키와 레드너가 밝혔다 [89]. 이로 부터 $\langle k_{nn} \rangle$ 은 다음과 같이 유도된다. 즉

$$\langle k_{nn} \rangle(k) = \sum_{k'} k' P(k'|k), \quad (32)$$

$$= \sum_{k'} \frac{k' P(k', k)}{k P(k)}, \quad (33)$$

$$\sim \sum_{k'} \frac{k' k'^{-(\gamma-1)} k^{-2}}{k^{1-\gamma}}, \quad (34)$$

$$\sim k^{-3+\gamma}. \quad (35)$$

여기서 $P(k'|k)$ 는 조건부 확률인데, 즉 어떤 한 연결선을 선택하였을 때, 그 중 한쪽 끝은 연결선수 k 를 가진 노드가 존재하는 조건으로 또 다른 쪽에 있는 노드의 연결선수가 k' 을 가질 확률이다. 위의 식에서 알 수 있는 바와 같이 $\gamma > 3$ 이 되면 성장하는 네트워크는 자연스럽게 유사성 네트워크가 된다.

X. 척도없는-네트워크의 여과현상

여과(percolation) 모형은 통계역학에서 잘 알려진 모형이다. 2 차원 사각격자에서 p 라는 확률로 링크(본드)가 존재하고 $1 - p$ 라는 확률로 존재하지 않는다고 하자. p 가 작으면 존재하는 링크의 개수가 적고, p 가 점점 커지면서 링크수가 증가하는데 임계확률 p_c 가 존재하여 $p > p_c$ 인 경우에는 계의 양쪽을 연결하는 여과송이가 생기게 된다. 따라서 $p < p_c$ 에는 여과송이가 없다가, $p > p_c$ 에는 여과송이가 확률 1로 생기는 상전이 현상이 생기게 되는데 이를 여과 전이(percolation transition)라고 한다. $p > p_c$ 에서의 여과송이의 크기는 $\mathcal{O}(N)$ 으로서 p 가 p_c 로부터 증가함에 따라 빠르게 증가한다. 임계확률에서는 어떤 노드가 s -크기의 송이에 속할 확률 $p(s)$ 는 멱함수 꼴을 따르게 된다. 즉 $p(s) \sim s^{-\tau}$.

앞에서 소개한 바와 같이 ER 모형은 N 개의 노드가 존재하는 정적 네트워크로, 두 노드 간에 연결선이 무작위

로 이어지게 된다. 총 연결선의 수가 $L > L_c \approx N/2$ 이 되면 계의 $\mathcal{O}(N)$ 개의 노드가 연결되는 여과송이가 형성된다. 이렇게 생겨난 여과송이에서 연결선수의 분포 함수는 푸아송 분포를 따른다. 그러면 여과송이에서 연결선수의 분포 함수가 멱함수를 따르는 척도없는-네트워크는 없을까? 앞서 소개한 스택틱 모형이 이러한 모형이라는 것이 자연스럽게 떠오를 수 있다. 왜냐하면 스택틱 모형에서 각 노드는 척도없는-네트워크가 되기 위해 필요한 조건을 처음부터 부여받았기 때문이다. 한편 L 이 매우 작은 경우에는 여과송이가 만들어질 수 없음은 당연한 것이다. 따라서, L 이 증가하면서 네트워크가 어떻게 변하며, 여과송이는 언제 형성되는가, 즉 여과전이이 일어나는 임계 연결선수 L_c 는 언제인가를 소개하고자 한다 [90].

L 이 매우 작은 값으로부터 증가함에 따라 노드들을 연결하는 작은 송이들이 형성되는데, 이런 송이들은 아직 성긴(sparse) 상태이며, 따라서 대부분의 송이들은 나뭇가지 구조(tree structure)를 가진다. 즉 각각의 송이들에는 아직 고리구조(loop structure)가 아직 형성되지 않았다. 나뭇가지 구조만 형성되는 경우에는 송이들의 구조적 특성을 갈래치기 과정(branching process) 방법 [91, 92]을 이용하여 해석적으로 이해할 수 있다. 이 방법을 이용하여 L 이 변하면서 송이들이 어떻게 진화하는지에 대한 연구가 최근 수행되었다 [90]. 이 연구 결과는 코헨 등 [93,94]이 바라바시-알버트 모형에서, 즉 연결선수가 $\mathcal{O}(N)$ 인 척도없는-네트워크에서 노드를 무작위로 제거하면서 언제 여과송이가 없어지는가를 살펴본 연구결과와 유사한 결과를 주기 때문에, 앞으로 소개할 내용은 척도없는-네트워크에서 일어나는 여과전이의 특징을 일반적으로 보여주는 결과라고 할 수 있다.

임계 연결선수 L_c 에서는 앞서 소개한 바와 같이 $p(s) \sim s^{-\tau}$ 이며 이로부터 $p(s)$ 의 모함수 $\mathcal{P}(z) = \sum_s p(s)z^s$ 는 특이성(singularity)을 갖게 된다.

$$\mathcal{P}(z) \sim b_1 + b_2(1-z)^{\tau-1} + \text{analytical terms.} \quad (36)$$

여기서 b_1 과 b_2 는 상수이다. 역으로 $p(s)$ 는 $\mathcal{P}(z)$ 를 알면 구할 수 있는데, $\mathcal{P}(z)$ 는 갈래치기 과정 방법을 이용하여 구할 수 있다. 이를 위해 $r(s)$ 이라는 양을 소개하자. 이 양은 s -크기의 송이가 매달려 있는 연결선의 수를 $2L$ 로 나눈 것이 된다. s -크기의 송이가 만들어지기 위해서는 어떤 한 노드 i 에 $s-1$ 크기의 송이가 매달려 있으면 된다. 만약 노드 i 의 연결선의 수 k_i 가 1 인 경우에는 $s-1$ 의 크기의 송이가 매달려 있으면 되고, 2 인 경우, 두 개의 연결선에 매달려 있는 송이의 크기의 합

이 $s-1$ 이면 된다. 이와 같은 것을 일반화하면 다음과 같은 식이 성립된다.

$$p(s) = P_d(0)\delta_{s,1} + \sum_{k=1}^{\infty} P_d(k) \sum_{s_1=1}^{\infty} \sum_{s_2=1}^{\infty} \cdots \sum_{s_k=1}^{\infty} \prod_{i=1}^k r(s_i) \delta_{\sum_{i=1}^k s_i, s-1}. \quad (37)$$

위 식을 간단히 하기 위해 $r(s)$ 의 모함수를 정의하자. $\mathcal{R}(z) = \sum_s r(s)z^s$. 식 (37) 에 모함수를 입히면,

$$\mathcal{P}(z) = zg(\mathcal{R}(z)) \quad (38)$$

이 성립되고, 여기서 $g(\omega) = \sum_{k=0}^{\infty} P_d(k)\omega^k$ 이다. 또한 이로부터,

$$\mathcal{R}(z) = zf(\mathcal{R}(z)), \quad (39)$$

이 성립된다. 여기서 $f(\omega) = g'(\omega)/\langle k \rangle$. 스택틱 모형에 대한 $P_d(k)$ 는 식 (29) 으로 알려져 있기 때문에 이것을 이용하면, $g(\omega)$ 를 구할 수 있다. 그 결과는 $1-\omega \ll 1$ 의 극한에서, $g(\omega) \simeq (1/N) \sum_{i=1}^N \exp[(k_i)(\omega-1)]$ 이며, 따라서, $\langle k_1 \rangle(1-\omega) \leq 1$ 의 경우에 $z = \omega + \sum_{n=1}^{\infty} a_n(1-\omega)^n$ 이다. 여기서 $a_1 = f'(1)$ 이며 a_2 는 음의 상수이다. 한편 $\langle k_1 \rangle(1-\omega) \gg 1$ 인 경우에는

$$z = \omega + \sum_{n=1}^{\lfloor \gamma-2 \rfloor} a_n(1-\omega)^n + A(1-\omega)^{\gamma-2} + \dots \quad (40)$$

인데 여기서, $\lfloor x \rfloor$ 는 floor 함수, 즉 $\lfloor x \rfloor$ 는 x 를 넘지 않는 가장 큰 정수이며, $A = \Gamma(3-\gamma)[\langle k_1 \rangle/N^{1/(\gamma-1)}]^{\gamma-2}$ 이다. 모함수 $\omega = \mathcal{R}(z)$ 는 $z = \omega/f(\omega)$ 의 역함수이며, $\mathcal{P}(z)$ 는 식 (38) 을 이용하여 구할 수 있다.

모함수 $\mathcal{P}(z)$ 의 임계점(critical point)은 $f'(1) = 1$ 의 조건으로부터 구할 수 있는데, 이것은 Molloy 등이 [95] 사용한 임계점의 조건식, $\langle k^2 \rangle / \langle k \rangle = 2$, 과 같다. 이 임계점으로부터 연결선수의 임계값 L_c 을 구할 수 있다.

$$L_c = \frac{1}{2} \frac{\zeta_N[1/(\gamma-1)]^2}{\zeta_N[2/(\gamma-1)]}. \quad (41)$$

여기서 $\zeta_N(x) = \sum_{i=1}^N i^{-x}$ 이다. L_c 는 γ 에 의존한다. 두 가지의 경우로 분류할 수 있는데, 즉 (I) $\gamma > 3$, (II) $2 < \gamma < 3$ 경우이다. N 이 매우 큰 극한에서 L_c 은 (I) 의 경우 $L_c \sim N$ 에 비례하고, (II) 의 경우에는 $L_c \sim N^{2(\gamma-2)/(\gamma-1)}$ 이다. 따라서, L_c/N 은 열역학 극한에서, 즉 $N \rightarrow \infty$ 의 극한에서, (I) 경우에는 유한한 값을 갖지만 (II) 경우에는 0 이 된다. 흥미로운 것은 N 이 유한한 경우에 $L > L_c(N)$ 에서 여과송이가 생기는 것이 통상

적인데 (I) 의 경우에는 가장 큰 송이의 크기가 $\mathcal{O}(N)$ 가 되지만, (II) 의 경우에는 비정상적인 현상이 벌어져 가장 큰 송이의 크기가 $\mathcal{O}(N^{(\gamma-2)/(\gamma-1)})$ 가 되어서 전체를 연결하지 못하게 된다.

유한한 송이들의 평균크기 $\langle s \rangle$ 를 생각해 보자. 여기서 ‘유한’하다는 것은 여과송이를 제외한다는 의미이다. 엄밀히 말하면 $L < L_c$ 인 경우에는 여과송이가 존재하지 않기 때문에 $\langle s \rangle = \sum_s s p(s)$ 이고 여기서 \sum_s 는 모든 송이들에 대한 합이다. L 이 커짐에 따라, 송이의 평균 크기는 점점 커지기 때문에 $\langle s \rangle$ 는 증가하게 된다. 한편 $L > L_c$ 인 경우에는 여과송이가 존재하기 때문에 \sum_s 는 여과송이 즉 가장 크기가 큰 송이를 제외한다. 그런데 이 영역에서는 L 이 커지면서 노드들이 여과송이에 참여할 확률이 점점 커지기 때문에 여과송이를 합에서 제외하면 $\langle s \rangle$ 는 L 이 커지면서 감소하게 된다. 따라서 $\langle s \rangle$ 은 $L = L_c$ 에서 봉우리 모양을 이루게 되는데 $N \rightarrow \infty$ 가 되면서 이 봉우리의 높이는 발산하게 된다. 이러한 현상은 유클리드 공간에서의 여과전이에서 일어나는 전형적인 성질이다. 그러나 척도없는-네트워크에서의 여과전이 현상에서는 $\gamma > 3$ 인 경우에는 봉우리가 L_c 에서 한 개 만들어지고 열역학 극한에서 발산하는 성질이 있지만 $2 < \gamma < 3$ 인 경우에는 봉우리가 L_c 외에도 $L > L_c$ 에서 하나가 더 생기는데 이 봉우리는 발산하지는 않는다. 따라서 $2 < \gamma < 3$ 의 척도없는-네트워크에서의 여과전이 현상은 정상적인 여과전이 현상과는 많은 차이가 있다. 자세한 내용은 최근 논문 [90]을 참고하기 바란다.

XI. 복잡계 네트워크의 분광분석

네트워크의 구조를 행렬로 표시하고 이 행렬의 분광적 성질을 살펴보는 것은 흥미로운 것이다. N 개의 노드가 존재하는 계를 $N \times N$ 의 행렬 \mathbf{A} 로 표시하고 만약 노드 i 와 j 가 연결되어 있으면 행렬 \mathbf{A} 의 (i, j) 과 (j, i) 의 성분, 즉 $a_{i,j} = a_{j,i} = 1$ 로 주어지고 아니면 0 인 이웃행렬(adjacency matrix)을 생각하자. 척도없는-네트워크의 행렬의 고유치 λ 의 분포는 그림 (3) 와 같다 [96, 97]. 이 고유치 분포함수는 중앙부분에서는 지수적으로 감소하지만, 꼬리부분에서는 멱함수적으로 감소하게 된다 (그림 3 참조). 최근의 이 분포함수에 대한 해석적인 해가 구하여졌고, 척도없는-네트워크의 경우 꼬리부분의 경우 $\rho(\lambda) \sim |\lambda|^{1-2\gamma}$ 이라는 것이 밝혀졌고, 이는 수치적인 결과와 일치하였다 [98].

네트워크 위에서의 막걸기(random walk) 문제를 생각하자. $u_i(t)$ 는 시각 t 에 노드 i 에 있는 막걸는 사람의 평

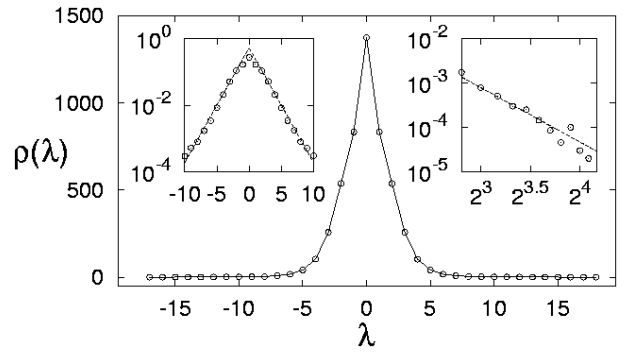


Fig. 3. Plot of the density of eigenvalues of the adjacency matrix \mathbf{A} versus eigenvalues for system size $N = 5,000$ averaged over 40 configurations for the BA model. Left inset: Semi-logarithmic plot of the density versus eigenvalues to show the exponential decay for small $|\lambda|$ s. Right inset: Double logarithmic plot of the density versus eigenvalues to show the power-law decay at spectrum edge.

균 수를 나타낸다. 그러면, 막걸기 동역학 현상은 다음과 같이 기술된다.

$$u_i(t+1) = \sum_j T_{ij} u_j(t). \quad (42)$$

여기서, T_{ij} 는 만약 i, j 가 연결되어 있으면 $T_{ij} = 1/k_j$ 이고 (k_j 는 노드 j 의 연결선수), 연결되어 있지 않으면 0 이다. 행렬 T 는 확산 행렬 Y 와 $Y = T - 1$ 의 관계를 가지고 있다. $t \rightarrow \infty$ 의 극한에서 노드 i 에 있는 막걸는 사람의 평균 수는 $u_i(\infty)$ 이 된다. 엄밀한 수학적 증명 없이 생각해 보면, 각 노드에 있는 막걸는 사람의 수가 그 노드의 연결선수에 비례한다고 하면 식 (42)은 자연스럽게 만족된다. 행렬 T 는 $T_{ij} \neq T_{ji}$ 이므로 대칭행렬이 아니다. 그러나 $\tilde{T}_{ij} = T_{ij} \sqrt{k_j/k_i}$ 의 관계로 행렬 \tilde{T} 를 정의하면 이 행렬은 대칭이 된다. 이 행렬은 원래의 행렬 T 와 같은 고유치를 가지며, 고유벡터는 $\tilde{v}_i^{(\alpha)} = v_i^{(\alpha)} / \sqrt{k_i}$ 의 관계를 갖는다. 여기서 $v_i^{(\alpha)}$ 는 행렬 T 의 (α) 번째 고유치를 갖는 고유벡터의 i 번째 성분이다. 행렬 \tilde{T} 의 고유치중 가장 큰 고유치의 고유벡터를 생각해 보면 $v_i^{(1)} \propto k_i$ 을 얻게 되는데, 이것은 $u_i(\infty)$ 에 비례하는 양이다. 따라서 막걸기 문제에서 장시간이 지난 후의 노드 i 에 있는 막걸는 사람의 밀도는 그 노드의 연결선수 k_i 에 비례한다 [99].

XII. 복잡계 네트워크의 견고성

척도없는-네트워크의 특징 중의 하나는 허브가 존재한다는 것이다. 허브의 존재로 인하여 네트워크의 견고

성에 많은 변화가 있다. 견고성을 측정하는 양으로 두 노드 간의 평균 거리, 즉 지름을 생각하자. 정과 공동연구자는 두 가지 방법을 통해 지름이 어떻게 변하는가를 살펴봐왔다 [101]. (i) 노드를 하나씩 무작위하게 선택하여 제거하면서 지름이 어떻게 변하는가를 측정하였다. (ii) 노드들 중에 가장 연결선수가 많은 노드 순으로 차례로 제거하면서 지름이 어떻게 변하는가에 대하여 연구하였다. (i) 과 (ii) 의 각각의 경우를 척도없는-네트워크와 무작위 네트워크에 대하여 수행하여 비교하였다. 정과 그의 공동연구자가 얻은 연구 결과는 (i) 의 경우에는 척도없는-네트워크가 무작위 네트워크에 비하여 견고하지만, (ii) 의 경우, 즉 의도적으로 외부에서 허브를 공격할 경우에는 척도없는-네트워크가 매우 취약한 반면, 무작위-네트워크는 견고한 면을 보여주었다. 이에 대한 해석적 연구는 갈래치기 과정 방법을 이용하여 연구되었다 [93,94,102]. 이러한 성질로부터 이해할 수 있는 것은 신진대사 반응 네트워크에서나 단백질 상호작용 네트워크 등에서 연결선수가 많은 노드들을 제거하는 경우 세포, 즉 생명체가 받는 타격이 매우 심각할 수 있다는 것이다. 이러한 성질을 이용하여 암세포 등의 신진대사 반응 네트워크의 허브를 제거함으로써 암세포를 사멸시킬 수 있는 방법을 생각해 볼 수 있다. 좀 더 구체적으로는 최근 E. Coli 박테리아의 신진대사 반응 네트워크에서 치명적 반응을 차단하여 생물 영양분이 만들어지지 못하도록 하는 연구가 수행되고 있다.

XIII. 복잡계 네트워크에서의 연쇄적 파멸 현상

척도없는-네트워크가 의도적인 공격에 매우 취약한 성질이 있는데 그것보다 더욱 심각한 피해는 도미노 현상에 의한 피해라고 할 수 있다. 척도없는-네트워크에서는 두 노드 간의 거리가 매우 좁기 때문에 한 노드의 피해는 다른 노드에 곧바로 영향을 끼친다. 따라서 도미노 현상이 유클리드안 네트워크에서보다 복잡계 네트워크에서 훨씬 잘 일어날 수 있다. 예를 들어 지난 8월 미국 동북부에서 일어난 전력망의 연쇄 정전 사태가 좋은 예라 할 수 있다 [103,104]. 전력망은 엄밀히 말하면 척도없는-네트워크가 아니다. 연결선수의 분포 함수를 조사해 보면 지수 함수를 따른다는 것이 알려져 있다. 그러나 척도없는-네트워크처럼 노드들 간의 연결이 매우 긴밀히 되어 있다. 전력망에서는 전력 수급을 효율적으로 수행하기 위해 한 지역에서 전력이 부족할 경우에 이웃에 있는 전력을 끌어와 쓰도록 설계되어 있다. 지난번 미국의 정전

사태는 이와 같은 복잡계 네트워크의 긴밀한 연결성 때문에 한 곳의 정전 사태가 동북부 전체의 정전사태로 삽시간에 파급되었다고 할 수 있다. 복잡계 네트워크에서의 연쇄적 파멸현상은 왓츠의 논문 [105]에서 강조되었다. 특히 복잡계 네트워크 상의 수송과정에서 일어나는 연쇄적 파국 현상에 대한 연구 또한 부분적으로 수행되었다 [106-109].

최근 복잡계 네트워크에서 일어나는 도미노 현상과 관련하여 사태현상의 중심적인 문제인 모래사태 모형인 Bak-Tang-Wiesenfeld (BTW) 모형을 척도없는-네트워크에 적용해 보는 시도가 있었다 [110]. BTW 의 모형은 다음과 같다. 척도없는-네트워크에서 노드 i 에 사태 문턱 z_i 를 부여하는데 그 문턱을 연결선수와 같도록 잡았다. 즉 $z_i = k_i$ 이다. 동역학적인 규칙은 다음과 같다. 임의의 노드 i 를 선택하여 그 노드에 모래를 하나 더한다. 이때 쌓인 모래의 높이를 h_i 라 하면 모래가 하나 더 쌓임에 따라 모래의 높이는 +1 만큼 증가하게 된다. 즉 $h_i \rightarrow h_i + 1$ 이 된다. 만약 모래의 높이가 그 노드의 사태 문턱보다 크거나 같게 될 경우에는 즉 $h_i \geq z_i$ 인 경우 즉 불안정한 상태에 놓이게 되면 그 높이를 z_i 만큼 줄여주고 즉 $h_i \rightarrow h_i - z_i$ 이 되고, 그 z_i 의 모래는 i 의 k_i 의 이웃에게 하나씩 나누어 주게 된다. 만약 노드 i 의 이웃 중의 하나인 노드 j 가 i 로부터 모래를 받아 $h_j \geq z_j$ 가 되는 경우에는 이곳에서도 앞의 경우와 같이 높이를 z_j 만큼 감소시키고 이웃에게 나누어 주는 동역학 작동을 실행하게 된다. 이와 같은 일련의 동역학 작동을 더 이상의 불안정한 노드가 없을 때까지 반복하게 된다. 더 이상의 불안정한 노드가 없는 경우에는 다시 임의의 노드를 선택하여 앞의 일련의 순서를 밟게 된다. 이러한 문제에서 관심을 갖게 되는 것은 외부로부터 모래가 주입되었을 때 연쇄적인 사태 과정을 통하여 불안정한 노드가 몇 개 생기는가 하는 것과 이에 대한 분포이다. 즉 사태 크기에 대한 분포함수이다. 이 분포함수를 $p_a(s)$ 라 하면 s 는 불안정한 노드의 개수를 의미한다.

척도없는-네트워크에서는 이러한 사태 크기 분포함수를 갈래치기 과정 방법(branching process approach)을 이용하여 구할 수 있다. 이를 위해 하나의 사태에 대하여 어떤 한 노드 i 가 k 개의 가지를 칠 확률을 $q_a(k)$ 라 하자. 여기서 k 는 노드 i 의 연결선수와 같다. $q_a(k)$ 는 두 가지 요소로 구성된다. $q_a(k) = q_1(k)q_2(k)$. 첫째로 $q_1(k)$ 는 노드 i 가 k 개의 이웃 노드로부터 모래 하나를 받을 확률을 의미한다. 둘째는 $q_2(k)$ 로, 이웃으로부터 하나의 모래를 받아 노드의 높이가 그 노드의 문턱과 동일하거나 그 이상이 될 확률이다. 즉 노드 i 의 높이 $k - 1$ 에 있을 확률이다. $q_1(k)$ 는 $q_1(k) = kP_d(k)/\langle k \rangle$ 이며, $q_2(k) = 1/k$

이다. 여기서 q_2 는 임의의 노드가 어떤 높이든지 동일한 확률을 가지고 존재한다는 가정하에서 얻은 결론이고, 이 가정은 수치적으로 검증되었다. 따라서, k 개의 갈래 치기가 이루어질 확률은

$$q_a(k) = q_1(k)q_2(k) = \frac{P_d(k)}{\langle k \rangle} \quad \text{for } k \geq 1, \quad (43)$$

로 주어지고 특별히 $k = 0$ 일 경우는 $q_a(0) = 1 - \sum_{k=1}^{\infty} q_a(k)$ 이다.

갈래치기 과정을 통하여 s 개의 노드가 불안정 상태가 되는, 즉 크기가 s 인 사태가 일어날 확률 $p_a(s)$ 는 $q_a(k)$ 로부터 다음과 같이 구할 수 있다 [91,92].

$$p_a(s) = q_a(0)\delta_{s,1} + \sum_{k=1}^{\infty} q_a(k) \sum_{s_1=1}^{\infty} \sum_{s_2=1}^{\infty} \cdots \sum_{s_k=1}^{\infty} \prod_{i=1}^k p_a(s_i) \delta_{\sum_{i=1}^k s_i, s-1}. \quad (44)$$

여기서 s_i 는 위의 k 개의 갈래 중에서 i 번째 갈래에서 파생되어 생기는 사태의 크기를 의미한다. 그런데 사태의 합이 s 이므로, $\sum_{i=1}^k s_i = s-1$ 이어야 한다. s 가 아니라 $s-1$ 인 이유는 갈래치기를 시작하는 노드가 배제되었기 때문이다. 식 (44) 로부터 $p_a(s)$ 를 구할 수 있는 방법은 모함수를 이용하는 것이다. 이를 위해 $p_a(s)$ 와 $q_a(k)$ 에 대한 모함수를 정의하자. $\mathcal{P}_a(y) = \sum_{s=1}^{\infty} p_a(s)y^s$ 와 $\mathcal{Q}_a(\omega) = \sum_{k=0}^{\infty} q_a(k)\omega^k$. 식 (44) 에 모함수를 적용하면

$$\mathcal{P}_a(y) = y \mathcal{Q}_a(\mathcal{P}_a(y)). \quad (45)$$

여기서 $\omega = \mathcal{P}_a(y)$ 라고 놓으면 이는 $y = \omega/\mathcal{Q}_a(\omega)$ 의 역함수가 된다. 그러나 이와 같은 관계식을 가지고만은 $p_a(s)$ 를 구하기 어렵다.

척도없는-네트워크에서 연결선수가 멱함수 꼴, $P_d(k) \sim k^{-\gamma}/\zeta(\gamma)$ 을 따른다고 하면, k 개의 갈래치기를 할 확률은

$$q_a(k) = \begin{cases} 1 - \zeta(\gamma)/\zeta(\gamma-1) & (k=0), \\ k^{-\gamma}/\zeta(\gamma-1) & (k \geq 1), \end{cases} \quad (46)$$

로 주어진다. 따라서, 이의 모함수 $\mathcal{Q}_a(\omega)$ 는 γ 에 의존하는 특이성을 가지게 되고 그 내용은 다음과 같다.

$$\mathcal{Q}_a(\omega) - \omega \simeq \begin{cases} A(\gamma)(1-\omega)^{\gamma-1}, & (2 < \gamma < 3), \\ -(1/2\zeta(2))(1-\omega)^2 \ln(1-\omega), & (\gamma = 3), \\ (1/2)B(\gamma)(1-\omega)^2, & (\gamma > 3). \end{cases} \quad (47)$$

여기서 우리들은 가장 주된 항만을 고려하였다. 여기서, $A(\gamma) = \Gamma(1-\gamma)/\zeta(\gamma-1)$, $B(\gamma) = [\zeta(\gamma-2)/\zeta(\gamma-1)]-1$

이다. 다음으로는 식 (45) 으로 부터 $\mathcal{P}_a(y)$ 를 구할 수 있고, 이로부터 사태 크기 분포함수 $p_a(s)$ 를 구할 수 있다.

$$p_a(s) \sim \begin{cases} a(\gamma) s^{-\gamma/(\gamma-1)} & (2 < \gamma < 3), \\ bs^{-3/2}(\ln s)^{-1/2} & (\gamma = 3), \\ c(\gamma) s^{-3/2} & (\gamma > 3). \end{cases} \quad (48)$$

여기서

$$\begin{aligned} a(\gamma) &= -A(\gamma)^{1/(1-\gamma)}/\Gamma[1/(1-\gamma)], \\ b &= \sqrt{\pi/6}, \\ c(\gamma) &= \sqrt{1/(2\pi B(\gamma))}. \end{aligned} \quad (49)$$

이러한 해석적인 결과는 $z_i = k_i^\eta$ ($\eta < 1$) 인 경우로 일반화할 수 있으며, 또한 모래를 뿌려주는 방법의 세부적 내용에는 무관함을 보일 수도 있다.

XIV. 복잡계 네트워크의 상전이 현상

복잡계 네트워크, 특별히 척도없는-네트워크 위에서 일어나는 상전이 현상을 살펴보기로 하자. 이 연구는 란다우 이론을 바탕으로 한 연구로서 이글로이와 터어반 [111], 도로고체프와 멘테스 [112,113]등이 수행하였는데, 임계현상을 규정하는 임계지수들이 네트워크의 연결선수 지수와 밀접하게 연관이 있다는 것이다. 중요한 결과는, 연결선수 지수가 어느 값 이상일 경우 임계현상이 표준적인 평균장 이론에서 관측되었던 것과 잘 들어맞으나, 연결선수 지수가 그보다 작을 경우에는 축척지수들이 연결선수 지수에 의존하여 연속적으로 변화한다는 것이다. 본 절에서는 이러한 결과를 살펴보도록 하겠다. 이는 척도없는-네트워크 앙상블에 대한 평균장이론으로 볼 수 있다.

우선 잘 알려진 q -상태의 팻츠모형의 하밀토니안을 생각해 보자.

$$-\frac{H}{k_B T} = \sum_{\langle ij \rangle} K_{ij} \delta(s_i, s_j) + \sum_i h_i \delta_{s_i, 0}. \quad (50)$$

여기서 노드 i 의 스핀 s_i 는 q 개의 상태, $s_i = 0, 1, \dots, q-1$ 를 갖는 팻츠 스핀이다. K_{ij} 은 i 와 j 의 상호작용인데 만약 두 노드간의 연결선이 있으면 $K_{ij} = K > 0$, 없으면 0 이다. h_i 는 외부에서 주는 $s_i = 0$ 방향으로의 자기장이다. 이에 대한 평균장 이론을 적용하면 다음과 같은 자체 모순없는 방정식을 얻게 된다.

$$\sum_j K_{ij} m_j = \sum_j K_{ij} \frac{\exp(\sum_\ell K_{j\ell} m_\ell + h_j) - 1}{\exp(\sum_\ell K_{j\ell} m_\ell + h_j) + q - 1}. \quad (51)$$

m_i 는 노드 i 의 자기화 값이다. 격자상수 z 를 가지고 있는 유클리디안 격자계의 경우, $m_i = m, h_i = h$ 라 놓으면, 위의 식은

$$m = \phi(zKm + h), \quad \phi(x) = \phi_{\text{Euclidean}}(x) = \frac{e^x - 1}{e^x + q - 1} \quad (52)$$

이 된다. $zKm + h$ 는 평균장 이론에 의한 유효마당이며, $\phi(0) = 0$ 이다.

일반적인 연결선수 분포 함수 $P_d(k)$ 를 갖는 척도없는-네트워크에서는 위의 자체 모순없는 방정식은 다음과 같이 쓰인다.

$$m = \sum_k P_d(k)k\phi(kKm + h)/\langle k \rangle. \quad (53)$$

이로부터 $h = 0$ 인 경우에 계의 자유에너지에 비례하는 식은 다음과 같이 쓸 수 있다.

$$f = \frac{\langle k \rangle K}{2} m^2 - \sum_k P_d(k)\Phi(kKm). \quad (54)$$

여기서 $\Phi(x)$ 는 위의 $\phi(x)$ 와 $d\Phi(x)/dx = \phi(x)$ 의 관계가 있다. 함수 $\Phi(x)$ 가 $x = 0$ 부근에서 해석적이라고 가정하면 위의 식은

$$f = \frac{\langle k \rangle K}{2} m^2 - \sum_{n=2}^{\infty} \frac{\Phi^{(n)}(0)}{n!} \langle k^n \rangle (Km)^n \quad (55)$$

와 같이 전개될 것으로 예측된다. 여기서 $\langle k^n \rangle = \sum_k P_d(k)k^n$ 인데 척도없는-네트워크의 경우 n 이 $[\gamma - 1]$ 보다 클 때 발산하게 된다. 식 (55)에서 $n \in [[\gamma], \infty)$ 인 발산하는 항들은 모두 더해져 결과적으로 해석적인 정규항들 외에 비해석적인 항을 만들게 된다.

$$f = \sum_{n=2}^{\infty} f_n m^n + Am^{\gamma-1}. \quad (56)$$

마지막 항의 계수 A 는 γ 에 의존하는 상수이며 지수 역시 γ 에 의존하는데 이 항으로 인해 자유에너지는 $m = 0$ 부근에서 더 이상 해석적이지 못하다. 참고로 지수 γ 가 정수일 경우에는 마지막 항에 로그함수 보정이 붙는다.

식 (56)을 통하여 임계점에서의 상전이현상을 기술하는 축척지수 값들을 구해보도록 하자. 질서매트 변수 m 값은 자유에너지를 최소로 만든다는 조건, 즉 $\partial f / \partial m = 0$ 으로부터 구해진다. 만약 연결선수 분포가 멱함수처럼 천천히 감소하지 않고 푸아송 분포나 규칙적 살창에서의 델타함수처럼 빨리 감소한다면 - 이는 사실 $\gamma \rightarrow \infty$ 에 해당한다 - 식 (56)은 m 이 작을때 근사적으로 $f \simeq f_2 m^2 + f_3 m^3 + f_4 m^4$ 와 같이 썬진다. 임계점 ($f_2 = 0$) 부근에서

질서매트 변수는 $f_2 > 0$ 일 때 $m = 0$ 이고, $f_2 < 0$ 일 때 $m \sim (-f_2)^\beta$ 이다. 여기서 임계지수 β 는 $f_3 > 0$ 일 경우 1 이며, $f_3 = 0$ 이고 $f_4 > 0$ 일 경우 1/2 이다. $q > 2$ 인 팻츠 모형의 경우처럼 $f_3 < 0$ 인 경우에는 불연속 상전이가 일어난다.

척도없는-네트워크의 경우를 보자. 주의해야 할 점은 비해석적 항인 $Am^{\gamma-1}$ 항이 자유에너지가 최소가 되는 m 값을 결정하는 데 참여하는 경우가 생긴다는 것이다. $f_3 > 0$ 인 경우 $3 < \gamma < \gamma_u = 4$ 일 때, 또는 $f_3 = 0, f_4 > 0$ 인 경우 $3 < \gamma < \gamma_u = 5$ 일때, 식 (56)은 근사적으로

$$f \simeq f_2 m^2 + Am^{\gamma-1} \quad (57)$$

으로 주어지고 임계지수 β 는 $1/(\gamma - 3)$ 이 된다. 각각 $\gamma > \gamma_u$ 일 때는 표준적인 평균장이론 또는 푸아송 연결선수 분포를 갖는 네트워크 위에서처럼 $\beta = 1$ 또는 1/2 이 된다. 참고로 γ 가 3보다 작은 경우에는 $\langle k^2 \rangle$ 이 발산하므로 상호작용 계수 K 가 유한하면 질서매트 변수 m 도 0이 아닌 유한한 값을 갖는다. 참고로 최근 척도없는-네트워크에서 q -상태 팻츠 모형에 대한 상전이에 대하여 좀 더 자세한 연구가 수행되었다 [113].

이상에서 살펴본 바와 같이 척도없는-네트워크 위에서 일어나는 상전이 현상의 임계지수는 연결선수 지수에 의존하여 연속적으로 변하는데 이는 매우 흥미로운 결과이다. 어떤 계의 상전이 현상을 이해하려 할 때 질서매트 변수의 임계지수가 보통의 평균장이론으로 얻은 값과 다를 경우, 그 계에 있는 강한 상호작용에서 비롯되었던 것에 반하여, 척도없는-네트워크에서는 연결선수의 불균일성 때문에 새로운 임계 현상이 일어나고 있는 것이 흥미롭다.

XV. 복잡계 네트워크에서의 수송 현상

수송 현상에 대한 이해는 네트워크에서 중요한 문제이다. 네트워크에서의 수송현상과 관련된 문제는 인터넷에서의 패킷 전달, 세포 내의 신호 전달, 사람들 사이의 정보 전달 등이 있을 수 있으며, 특히 요즘 인터넷에서의 정보사냥이 정보산업의 핵심적인 역할을 하므로 이에 대한 네트워크에서의 수송현상을 연구하는 것은 매우 중요하다고 할 수 있다. 네트워크에서의 수송현상은 다공물질처럼 무질서계에서 일어나는 확산현상이 아니라 가장 빠르게 이동하려는 의도가 있기 때문에 단순한 확산운동을 연구하는 것은 아니다.

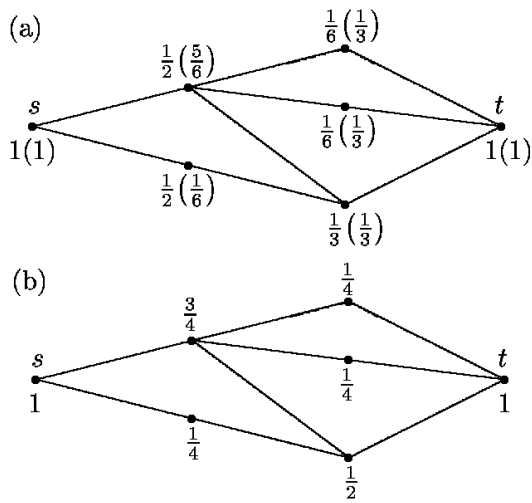


Fig. 4. The definitions of the load (a) and the BC (b). In (a), the load at each vertex due to a unit packet transfer from the vertex s to the vertex t , $\ell^{(s \rightarrow t)}$ is shown. In the parentheses, $\ell^{(t \rightarrow s)}$ is also shown for the packet transport from t to s for comparison. In (b), the BC at each vertex due to a packet transfer between s and t , $b^{(s,t)}$, is shown. The BC $b^{(s,t)}$ is equal to $b^{(t,s)}$.

1. 최단 경로 수송과 부하

예를 들어 A 라는 사람이 B 라는 사람에게 어떤 정보를 전달하려고 한다고 하자. A와 B 사이에 직접적인 인연이 없을 경우 가장 효율적인 방법은 A와 B를 연결하는 사회연결망 위에서 최단 경로를 통해 정보를 전달하는 것일 것이다. 이렇게 “의도된 수송”의 경우 일반적으로 “최단 경로 수송”을 생각할 수 있다. 여기서 최단 경로란 A 에서 B 까지를 연결하는 최소 링크 수를 갖는 길(path)로 정의한다. 물론 실제 상황에서는 문제가 그렇게 간단하지 않다. 사람 간의 관계에는 친밀도, 신용 등 여러 변수가 있고, 인터넷 상의 정보 전달에도 링크의 대역폭 등을 고려할 때 최단 경로라는 개념이 적절하지 않은 경우가 있을 수가 있다. 이러한 문제는 가중 네트워크(weighted networks)에서의 수송현상을 생각하여야 하는데 이에 대한 연구는 아직 이루어지지 않고 있다. 이 절에서는 가중치가 없는 네트워크에서 모든 노드들이 서로 정보를 주고 받는 경우만을 생각하기로 하자.

모든 노드들은 다른 모든 노드로 단위 시간당 한 단위의 정보를 전달하고, 이들의 모든 정보는 노드들 간의 최단 경로를 통해 전달된다. 우리가 던지는 질문은 이러한 수송 과정에서 각각의 노드들이 얼마만큼의 역할을 할 것이냐, 또는 얼마나 많은 정보를 전달해야 하는냐 하는 것이다. 부하(load)라 부르는 이 양은 어떤 노드가 모든 노드쌍 사이의 최단 거리에 얼마나 많이 참여하는가 하

는 위상적인(topological) 양으로 해석될 수도 있다 [71]. 또한 부하는 이미 사회학에서 1970년대에 제안된 사이 중앙성(betweenness centrality)과 밀접하게 관계된 양이다 [114]. 구체적으로 살펴보면 노드 s 에서 t 로 단위 데이터를 보내는 경우 Fig. (4a)에서 보는 바와 같이 두 노드간의 최단거리를 따라 보낸다고 하자. 데이터가 이동하면서 경로가 나뉠 때 데이터가 똑같이 나뉜다고 하면 노드 k 를 통해 지나간 데이터의 양 $\ell_{s \rightarrow t}^{(k)}$ 은 Fig. (4a)에서 명시한 것과 같다. 이러한 정의에 의하면 s 에서 t 로 보내지는 경우와 t 에서 s 로 보내지는 경우에 각 노드를 지나간 데이터의 양이 다르게 된다. 부하란 모든 노드에서 다른 노드로 단위 데이터를 최단거리를 통해 보낼 때 각 노드를 통해 지나간 데이터의 양을 표시한 것이므로 $\ell_k = \sum_{s,t} \ell_{s \rightarrow t}^{(k)}$ 이 된다. 중앙성의 값은 부하와는 약간 다르게 정의된다. Fig. (4b)에서 볼 수 있는 바와 같이 우선 $s \rightarrow t$ 또는 $t \rightarrow s$ 로 가는 경로 수를 세면, (이 경우에는 그 두 양이 서로 같다) 각 노드에 걸리는 중앙값은 s 에서 t 로 데이터를 보낼 때 k 라는 노드를 지나가는 최단 경로가 몇 개 있는가를 의미한다. 만약 데이터가 더 이상 쪼개질 수 없는 단위로 나뉘었고 $s \rightarrow t$ 의 최단 경로가 3개 존재할 때, $s \rightarrow t$ 로 9개의 단위 데이터를 보냈을 경우 각 경로를 통해 3개의 데이터가 지나갈 수 있다는 의미이다. 임의의 한 노드에서의 부하량과 중앙성 값은 위의 예처럼 같지 않을 것 같지만, 가능한 모든 노드 쌍에 대하여 계산하기 때문에 두 양은 통계적으로 구별할 수 없는 거의 같은 값을 갖기 때문에 사람들이 구별 없이 쓰는 경우가 많다. 중앙성값을 구하는 빠른 알고리즘이 개발되어서 학계에서 널리 사용되고 있다 [115,116].

병진 불변성을 갖는 유클리디안 공간에서나 통계적 균일성을 갖는 무작위 네트워크에서와 달리 복잡계 네트워크에서는 노드들간의 연결선수의 불균일성으로 인해 각 노드의 부하에도 불균일성이 나타난다. 연결선수가 노드와 이웃 노드만을 고려하는 국소적인 양임에 반해 부하는 전체적인 정보를 반영하는 양으로, 많은 경우 그 불균일성이 증폭되어 나타난다. 특히 척도없는-네트워크의 경우 부하의 분포함수도 멱함수 법칙을 따른다는 것이 알려졌다. 즉 $P_L(\ell) \sim \ell^{-\delta}$ 이다. 또한, 실재하는 또는 모형상의 여러 척도없는-네트워크들의 분석을 통해 부하 분포함수의 지수가 연결선수 분포함수의 변화에 대해서 많이 변하지 않는다는 것이 알려졌다. (그림 (5) 참조). 나아가 부하 분포함수의 지수가 척도없는-네트워크의 보편양(universal quantity)일 수 있다는 추측까지 낳았다. 이 문제는 아직 해석적인 뒷받침을 받지 못해 미해결의 상태로 남아있다 [117,118].

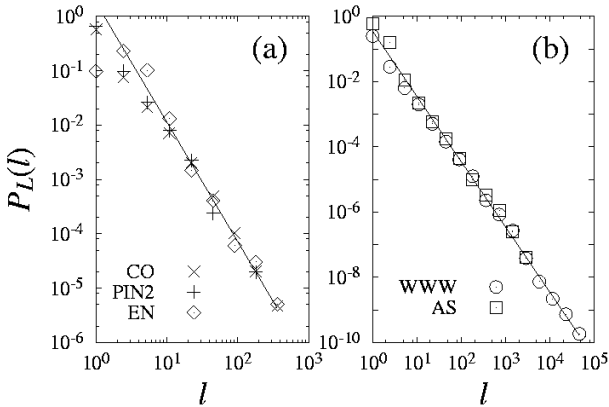


Fig. 5. The load distributions for the coauthorship network, the yeast protein interaction network, and the metabolic network of eukaryotes in (a), and the world-wide web and the Internet at the autonomous system level in (b). The slope $\delta \approx 2.2(1)$ in (a) and $\delta \approx 2.0$ in (b).

2. 부하 분포와 척도없는-네트워크의 위상적 성질

나무(tree)는 고리(loop)가 하나도 없는 네트워크이다. 고리가 있고 없음이 수송 현상, 또는 부하에 미칠 영향을 생각해 보자. 고리가 없는 나무의 경우 모든 노드들간의 최단 경로는 유일하다. 그러나 고리가 있는 경우, 일반적으로 최단 경로는 유일하지 않다. 고리구조의 존재는 그래프이론에서 항상 중요한 문제이고 최근 척도없는-네트워크에서 고리구조에 대한 관심을 나타낸 논문도 있다 [119]. 고리구조로 인하여 단순한 삼각형을 넘어 지름길을 제공하는 긴 고리가 서로 얽혀있는 경우 고리들은 꼭 묶인 속쇄(core)를 형성한다. 이러한 네트워크의 위상적 차이는 첫째, 부하 분포함수의 지수에 반영된다. 바라바시-알버트 모형과 같이 성장하는 척도없는-나무 네트워크의 경우 연결선수 지수에 관계없이 부하 지수가 $\delta = 2$ 임이 해석적으로 알려져 있다. 반면에, 많은 실제 또는 모형상의 척도없는-네트워크에서는 부하 분포함수 지수가 약 $\delta \approx 2.2$ 가 됨이 관측되었다. 둘째, 소위 질량-거리 관계식에도 반영된다. 질량-거리 관계식이란 노드쌍 간의 최단 경로의 거리 d 와 그 최단 경로상에 자리하는 노드의 수 $M(d)$ 간의 관계식이다. (그림 (6) 참조). 최단 경로가 유일하다면 두 양은 같고 $M(d) = d$ 의 관계식을 보이겠지만 실제로는 고리의 존재로 인해 둘의 관계는 자명하지 않게 된다. 특히 고리들이 속쇄 또는 덩어리(blob)를 형성하는 경우 질량-거리 관계식은 혹(hump)을 보인다. 반면 부하 분포함수의 지수가 2인 경우 둘은 선형관계를 보인다 [120].

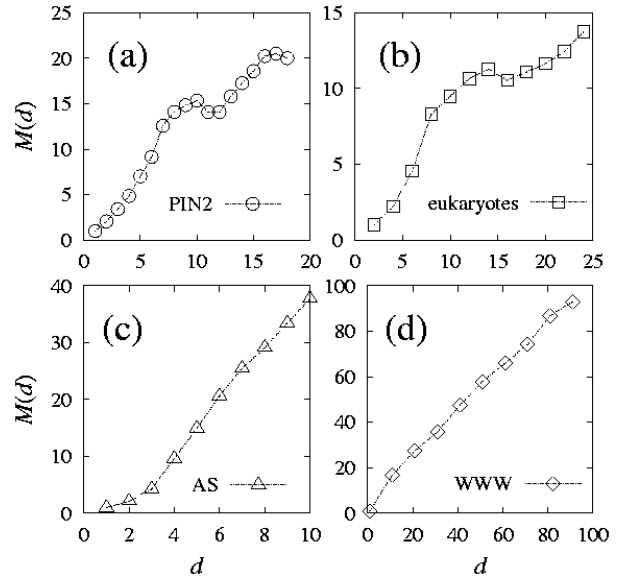


Fig. 6. The mass-distance relation for prototypical SF networks, the yeast protein interaction network (a), the metabolic network of eukaryote (b), the autonomous system (c), and the world-wide web.

3. 부하간의 상관관계

연결선수의 상관관계를 기술하기 위해 도입된 유사성 계수와 비슷하게 임의의 한 연결선을 무작위로 선택하여 양쪽의 노드에 걸리는 부하간의 상관관계를 생각해 보자. 이를 위해 앞에서 정의한 k 라는 연결선수를 가지고 있는 노드의 이웃 노드의 평균 연결선수 $\langle k_{nn} \rangle(k)$ 와 비슷하게 ℓ 이라는 부하를 가지고 있는 노드의 이웃 노드들의 평균 부하 $\langle \ell_{nn} \rangle(\ell)$ 를 Fig. (7)에서 조사할 수 있다. 연구 결과 생물학적 네트워크나 정보 네트워크에서는 연결선수 유사성 계수와 부하 유사성 계수가 비슷하게 나타난 반면 사회 네트워크에서는 부하 유사성 계수는 연결선수 유사성 계수의 거의 10분의 1로 거의 0에 가깝게 나타났다. 이러한 낮아진 상관 관계는 양의 연결선수 유사성 계수의 결과이며 사회학적으로는 모든 개인이 그 역할에 관계없이 평균적으로 비슷한 중요도를 지닌 사람들과 이웃한다는 의미이다. 즉 개인의 영향력 정도에 관계없이, 한 사람의 주위에는 영향력이 많은 사람, 적은 사람들이 항상 공존한다는 의미이다 [121]. 대통령 주위 사람이나 평시민의 주위 사람이나 영향력의 다소는 비슷하다는 의미로 해석된다. 이를 확인하기 위해 부하와 연결선수와의 관계를 그림 (8)에서 조사하였다. 여기서 비유사(dissortative) 또는 중립(neutral) 네트워크는 부하와 연결선수가 비슷하게 자리매김하여 $\ell \sim k^{(\gamma-1)/(\delta-1)}$ 의 관계를 보이지만 유사 assortative) 네트워크는 ℓ 과 k 가 일정한 관계가 없음을 볼 수 있다.

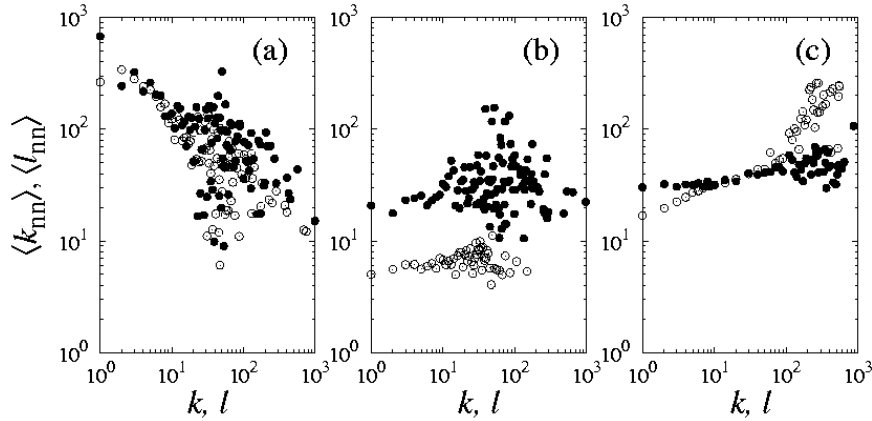


Fig. 7. Plot of $\langle k_{nn} \rangle(k)$ (○) and $\langle l_{nn} \rangle(l)$ (●) for (a) the Internet on the level of autonomous systems (dissortative), (b) the non-degenerate configuration model with $\gamma = 3$ (neutral), and (c) the coauthorship network in the field of neuroscience (assortative). All data are obtained from a single configuration.

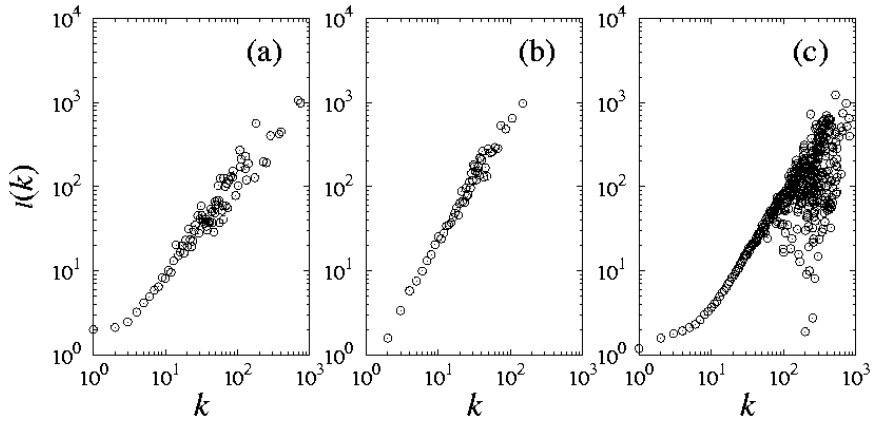


Fig. 8. Plot of the degree-load relation for (a) the Internet on the level of autonomous systems (dissortative), (b) the non-degenerate configuration model with $\gamma = 3$ (neutral), and (c) the coauthorship network in the field of neuroscience (assortative).

4. 확산 수송과 막걸기 중앙성

좀 더 일반적으로 네트워크에서의 수송현상이 최단 경로 이외의 경로를 따르는 경우에 대하여 생각해 보자. 이 절에서는 확산현상에 대하여 생각해 보자. 네트워크 위에서의 일반적인 확산 운동은 좁은세상-네트워크에 대해서 모나송 [122]에 의해, 척도없는-네트워크에 대해서는 노와 리거 [123]에 의해 연구되었다. 그러나 확산 수송의 경우 우리는, 갑에서 출발하고 을에서 흡수되는, 경계가 주어진 흡수 확산을 생각한다. 앞의 최단 경로 수송에서와 같이 이러한 확산 수송에 대하여 각각의 노드의 역할을 정의할 수 있고 이러한 양을 막걸기 중앙성이라 부른다 [124]. 막걸기 중앙성의 분포는 연결선수 분포와 같은 분포를 따르는데, 이는 막걸기 중앙성이 연결선수에 비례하기 때문이다.

XVI. 네트워크에서의 집단구조의 분류

실재하는 네트워크는 많은 경우 집단구조(modular structure)를 가지고 있다. 인터넷과 같은 경우에는 지역적으로 집단구조가 형성되고, 단백질상호작용 네트워크에서는 기능적으로 집단구조를 가지고 있다. 또 논문 공저자 네트워크와 같은 사회연결망에서도 사람들의 유유상종성 때문에 자연스럽게 집단구조가 형성된다. 이와 같은 집단구조가 형성되어 있는 네트워크에서 어떻게 집단을 구별해낼 수 있을까? 이러한 집단분류에 대한 효율적 방법에 대한 연구가 최근 활발히 일어나고 있다.

집단분류 방법의 하나로 뉴우만은 위에서 소개한 부하라는 것을 이용하여 분류작업을 수행하였다. 이에 대한 방법은 다음과 같다 [125]. (i) 네트워크에서 모든 노드와 링크에 대한 부하를 계산한다. 여기서 링크 부하는

앞 절에서 소개한 부하를 링크에 자연스럽게 적용한 개념으로서 노드를 통해서 지나간 데이터의 양 대신 링크를 지나간 데이터의 양으로 정의한다. (ii) 모든 링크 중에서 부하가 제일 많은 링크를 선택하여 제거한다. (iii) 제거된 링크가 있는 네트워크에서 각 링크에 걸리는 부하를 계산하고, 다시 부하가 제일 많은 링크를 제거한다. 이와 같은 계산을 반복하면, 집단과 집단을 연결하는 링크가 부하가 크기 때문에 제거되고 집단을 형성하는 송이들이 노출되게 된다. 이와같은 방법은 집단구조를 추출하는데 매우 효과적인 방법이 될 수 있으나 계산과정에 많은 시간이 걸리기 때문에 노드 수가 많은 네트워크에서는 사용하기 힘들게 된다. 또 언제 링크를 제거하는 절차를 중단하여야 하는가에 대하여 명확한 제시가 없다. 참고로 네트워크의 집단구조를 분류해내는 다른 방법도 소개되었는데 다음 참고문헌을 참고하기 바란다 [126-128].

XVII. 척도-없는 네트워크의 복잡성과 지름 변화

아주 조그마한 섭동이 주어진 경우에 복잡계는 다양한 반응을 보여서 그 반응의 정도를 확실히 예측할 수 없다는 것이 복잡계의 큰 특징이라고 할 수 있다. 예를 들면, 대표적인 물리 복잡계인 스핀유리(spin glass)의 경우, 외부에서 걸어주는 자기장의 세기나 방향이 아주 조금만 변해도 스핀의 방향이 집단적으로 변하게 되는 Barkhausen 잡음(Barkhausen noise)을 보인다. 이 때 방향이 바뀐 스핀의 수는 매우 다양한 양상을 보이고 확실히 예측할 수 없다. 물리계 외의 대표적인 복잡계인 주식 시장의 경우, 외부 상황이 변함에 따라 주식 가격이 변화하는데, 그 변화량의 분포가 멱함수법칙을 따른다는 것이 알려져 있다. 이는 주식 가격의 변화를 누구도 확실히 예상할 수 없다는 사실을 확인한다. 그럼에도 불구하고, 이렇게 복잡계의 반응을 확실히 예측할 수는 없으나 특정 반응을 보일 확률은 이야기할 수 있다. 즉, 예측 불가능성이 확률적인 예측으로 대체될 수 있다는 것이다. 복잡계 네트워크에서는 가장 간단하고 중요한 양중 하나가 두 노드간의 평균거리인데 최근 연구에서는 작은 섭동에 의한 상태 변화의 간단한 예로 척도없는-네트워크의 지름이 단일 노드 제거를 통해 얼마나 변화할지를 살펴보았다. 그 결과 약 96%에 이르는 대부분의 경우 지름의 변화가 0.02% 미만이었지만 나머지 약 4%의 경우 그 영향이 매우 커서 지름 변화들 $\{\Delta_i \equiv (D_i - D_0)/D_0\}$ (여기서 D_0 는 섭동이 가해지지 않은 상태의 지름이고,

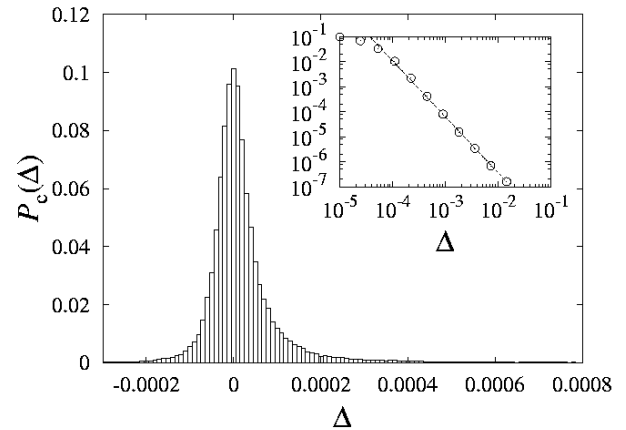


Fig. 9. Normalized histogram of the diameter changes for the static model with $\gamma = 3$ and $N = 10^4$, averaged over 10 configurations. Horizontal range is truncated for clearance, but runs up to 2×10^{-2} . Inset: Plot of $P_c(\Delta)$ in log-log scale for $\Delta > 0$. Dashed line is a fit line having a slope -2.2 . Data points are logarithmically binned.

D_i 는 노드 i 가 제거된 상태에서의 지름이다) 의 분포 함수 $P_c(\Delta)$ 는 멱함수 법칙의 꼬리를 형성하였다. 이는 척도없는-네트워크 역시 자기조직화를 통해 고비 상태에 있다는 가능성과 더불어 단일 노드 제거에 의한 네트워크 지름 변화의 예언(prediction)이 무의미함을 암시한다. 또한 이는 네트워크의 선형응답이론의 출발점이 될 수 있는 결과로서 앞으로 이러한 방향의 연구가 많이 요구된다.

XVIII. 가중치 복잡계 네트워크

지금까지는 연결선이 가지고 있는 가중치를 고려하지 않고 연결선이 있으면 1, 없으면 0 의 세기를 갖는 경우만을 생각하였다. 그런데 실재하는 네트워크에서는 연결선의 가중치가 있는 경우가 종종 있다. 즉 가중 네트워크를 생각하자. 이러한 경우에 네트워크의 지름이 연결선 가중치의 분포에 따라 달라질 수 있다. 네트워크 상에서 각 연결선에는 그 연결선을 통과하면서 들어가는 비용이 있으며, 두 노드를 연결하는 길은 그 길을 따라 치러야 하는 총비용이 가장 적은 길(최소비용길)을 택하게 된다. 가중치의 분포에 따라, 강한 무질서의 경우와 약한 무질서의 경우가 있는데, 강한 무질서의 경우는 임의의 두 노드를 이어주는 최소 비용길을 구성하는 연결선 중에서 가장 비용이 많이 들어가는 연결선의 비용이 총 비용과 거의 같은 경우를 말한다. 강한 무질서가 있는 경우, ER 이나 WS의 경우, 지름은 $D \sim N^{1/3}$ 에 비

레하고, 척도없는-네트워크의 경우에는 연결선수의 지수가 $3 < \gamma < 4$ 인 경우에는 $D \sim N^{(\gamma-3)/(\gamma-1)}$ 이며, $2 < \gamma < 3$ 인 경우에는 $D \sim \ln^{\gamma-1} N$ 이다. 약한 무질서가 있는 경우에, ER, WS, 이나 척도없는-네트워크 모두 $D \sim \ln N$ 이다 [130].

XIX. 끝내기

지금까지 복잡계 네트워크에 대하여 물리학계에서 연구한 내용을 소개하였는데, 본 논문에서 소개된 내용은 전체 논문들의 극히 일부라고 생각하여도 된다. 생략된 중요한 내용들은 전염현상 [131,132] 과 면역현상 [133-135], 동기화 문제 [136-139], 정보사냥 문제 [140-143], 부울리안 네트워크 [144], 생물 네트워크 [145,146], 경제 네트워크 [147-149], 신경망 [150] 연구 등이 있다. 이와 같은 문제들은 아직 해결하여야 할 문제들이 많이 있다. 또 너무 많은 논문들이 쏟아져 나오고 있기 때문에 참고 문헌에서 중요한 논문들이 많이 생략된 점을 이해하여 주길 바란다.

감사의 글

이 논문은 과학재단의 선도기초과학연구실 (ABRL) 지원 연구비 R14-2002-059-010000-0 에 의해 수행되었음.

참고 문헌

- [1] S. Milgram, *Psychology Today* **2**, 60 (1967).
- [2] P. Erdős and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960); *Bull. Inst. Int. Stat.* **38**, 343 (1961).
- [3] S. Wasserman and K. Faust, *Social Network Analysis*, (Cambridge University Press, Cambridge, 1994).
- [4] Z. N. Oltvai and A.-L. Barabási, *Science* **298**, 763 (2002).
- [5] B.A. Huberman, *The Laws of the Web* (MIT Press, Cambridge, MA, 2001).
- [6] D. J. Watts and S. H. Strogatz, *Nature (London)* **393**, 440 (1998).
- [7] R. Albert, H. Jeong and A.-L. Barabási, *Nature* **401**, 130 (1999).
- [8] B. A. Huberman and L. A. Adamic, *Nature* **401**, 131 (1999).
- [9] A. L. Barabási and R. Albert and H. Jeong, *Physica A* **272**, 173 (1999).
- [10] M. E. J. Newman, *J. Stat. Phys.* **101**, 819 (2000).
- [11] K. Ziemelis and L. Allen, *Nature (London)* **410**, 241 (2001) and following review articles on complex systems.
- [12] R. Gallagher and T. Appenzeller, *Science* **284**, 87 (1999) and following viewpoint articles on complex systems.
- [13] G. Parisi, *Physica* **263A**, 557 (1999).
- [14] A.-L. Barabási, *Linked: The New Science of Networks* (Perseus Press, New York, 2002).
- [15] M. Buchanan, *Nexus: Small Worlds and the Groundbreaking Science of Networks* (W.W. Norton & Co, New York, 2002).
- [16] D. J. Watts, *Six Degrees: The Science of a Connected Age* (New York: Norton, 2003).
- [17] S.A. Kauffman, *The origin of order: Self-organization and selection in evolution* (Oxford University Press, New York, Oxford, 1993).
- [18] S. Bornholdt and H. G. Schuster, ed., *Handbook of Graphs and Networks: From the Genome to the Internet* (Germany: Wiley-VCH, 2003).
- [19] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (New York: Oxford University Press, 2003).
- [20] B. Bollobás, 2nd Ed. *Random Graphs* (England: Cambridge University Press, 2001).
- [21] P. Baldi, P. Frasca and P. Smyth, *Modeling the Internet and the Web: Probabilistic Methods and Algorithms* (England: John Wiley & Sons, 2003).
- [22] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **47**, 74 (2002).
- [23] S. N. Dorogovtsev and J. F. F. Mendes, *Adv. Phys.* **51**, 1079 (2002).
- [24] M. E. J. Newman, *SIAM Review* **45**, 167 (2003).
- [25] S.N. Dorogovtsev and J.F.F. Mendes, *Phys. Rev. E* **63**, 025101(R) (2001).
- [26] V. Pareto, *Cours d'Economie Politique*, vol.2 (Université de Lausanne, Lausanne, 1897).
- [27] B.C. Arnold, *Pareto distribution* (Encyclopedia of Statistical Sciences, V.6, John Wiley & Sons, 1985).

- [28] E.V. Koonin, Y.I. Wolf and G.P. Karev, Nature (London) **420**, 218 (2002).
- [29] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabasi, Science **297**, 1551 (2002).
- [30] R. Pastor-Satorras, A. Vázquez and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
- [31] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
- [32] B. Bollobás, Discrete Math. **33**, 1 (1981).
- [33] F. Chung and L. Lu, Adv. Appl. Math. **26**, 257 (2001).
- [34] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman and S. H. Strogatz, Phys. Rev. E **64**, 041902 (2001).
- [35] S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, Phys. Rev. E **64**, 066110 (2001).
- [36] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins and E. Upfal, *The Web as a graph*, Proceedings of the 19th ACM Symposium on Principles of Database Systems, pp 1-10 (2000).
- [37] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins and E. Upfal, *Stochastic models for the web graph*, Proceedings of the 41th IEEE Symposium on Principles on Foundations of Computer Science.
- [38] R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins, *Extracting large-scale knowledge bases from the web*, Proceedings of the 25th VLDB Conference, Edinburg, p639 (1999).
- [39] R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins, *Trawling the Web for emerging cyber-communities*, <http://www.almaden.ibm.com/cs/k53/trawling.ps> of Database Systems, pp 1-10 (1999).
- [40] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins and J. Wiener, *Graph structure in the web*, Comput. Netw. **33**, 309 (2000).
- [41] M. Faloutsos, P. Faloutsos and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
- [42] K.-I. Goh, B. Kahng and D. Kim, Phys. Rev. Lett. **88**, 108701 (2002).
- [43] E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).
- [44] IMDb.com.
- [45] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
- [46] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
- [47] C. Tsallis and M.P. de Albuquerque, Eur. Phys. J. B **13**, 777 (2000).
- [48] A. Vázquez, Preprint (cond-mat/0105031) (2001).
- [49] S. A. Kauffman, *The origin of order: Self-organization and selection in evolution* (Oxford University Press, New York, Oxford).
- [50] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai and A.-L. Barabási, Nature (London) **407**, 651 (2000).
- [51] A. Wagner and D. A. Fell, Proc. R. Soc. London B **268**, 1803 (2001).
- [52] H.-W. Ma and A.-P. Zeng, Bioinformatics **19**, 1423 (2003).
- [53] J. C. Venter *et al.*, Science **291**, 1304 (2001).
- [54] E. F. Keller, *The Century of the Gene* (Cambridge, Mass.: Harvard University Press, 2000).
- [55] S. Fields and O. Song, *A Novel Genetic System to Detect Protein-Protein Interactions*, Nature **340**, 245 (1989).
- [56] L. Zhu and G. J. Hannon, eds., *Yeast Hybrid Methods* (Natick, MA: Eaton, 2000).
- [57] P. Uetz *et al.* Nature (London) **403**, 623 (2000).
- [58] T. Ito, K. Tashiro, S. Muta, R. Ozawa, T. Chiba, M. Nishizawa, K. Yamamoto, S. Kuhara, and Y. Sakaki, Proc. Nat'l. Acad. Sci. USA **97**, 1143 (2000).
- [59] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, Proc. Nat'l. Acad. Sci. USA **98**, 4569 (2001).
- [60] I. Xenarios, L. Salwinski, X. J. Duan, P. Higney, S.-M. Kim and D. Eisenberg, Nucleic Acids Res. **30**, 303 (2002).
- [61] H. W. Mewes, D. Frishman, U. Guldener, G. Mannhaupt, K. Mayer, M. Mokrejs, B. Morgenstern, M. Munsterkotter, S. Rudd and B. Weil, Nucleic Acids Res. **30**, 31 (2002).
- [62] H. Jeong, S. P. Mason, A.-L. Barabási and Z. N. Oltvai, Nature (London) **411**, 41 (2001).
- [63] A. Wagner, Mol. Biol. Evol. **18**, 1283 (2001).
- [64] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
- [65] S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).

- [66] P. L. Krapivsky and S. Redner, *J. Phys. A* **35**, 9517 (2002).
- [67] P. L. Krapivsky, S. Redner and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).
- [68] G. Bianconi and A.-L. Barabási, *Europhys. Lett.* **54**, 439 (2001).
- [69] G. Bianconi and A.-L. Barabási, *Phys. Rev. Lett.* **86**, 5632 (2001).
- [70] C.W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1983).
- [71] K.-I. Goh, B. Kahng and D. Kim, *Phys. Rev. Lett.* **87**, 278701 (2001).
- [72] G. Caldarelli, A. Capocci, P. De Los Rios and M.A. Muñoz, *Phys. Rev. Lett.* **89**, 258702 (2002).
- [73] R. Albert and A.-L. Barabási, *Phys. Rev. Lett.* **85**, 5234 (2000).
- [74] R.V. Solé, R. Pastor-Satorras, E.D. Smith and T. Kepler, *Adv. Compl. Syst.* **5**, 43 (2002).
- [75] J. Kim, P. L. Krapivsky, B. Kahng and S. Redner, *Phys. Rev. E*, **66**, 055101 (2002).
- [76] A. Vázquez, A. Flammini, A. Maritan and A. Vespignani, *ComplexUs* **1**, 38 (2003).
- [77] J. Berg, M. Lässig and A. Wagner, e-print (<http://arxiv.org/abs/cond-mat/0207711>) (2002).
- [78] G. P. Karev, Y. I. Wolf, A. Y. Rzhetsky, F. S. Berzovskaya and E. V. Koonin, *BMC Evol. Biol.* **14**, 18 (2002).
- [79] S. Wuchty, *Mol. Biol. Evol.* **18**, 1694 (2001).
- [80] J. Park, M. Lappe and S.A. Teichmann, *J. Mol. Biol.* **307**, 929 (2001).
- [81] K.-I. Goh, B. Kahng and D. Kim (arXiv:q-bio.MN/0312009).
- [82] E. M. Jin, M. Girvan and M. E. J. Newman, *Phys. Rev. E* **64**, 046132 (2001).
- [83] D. J. Watts, P. S. Dodds and M. E. J. Newman, *Science* **296**, 1302 (2002); *Proc. Natl. Acad. Sci. USA* **99**, 5766 (2002).
- [84] J. M. Kleinberg, in T. G. Dietterich, S. Becker and Z. Ghahramani (eds.), *Proceedings of the 2001 Neural Information Processing Systems Conference*, (MIT Press, Cambridge, MA, 2002).
- [85] D.-H. Kim, B. Kahng and D. Kim, to appear in *Eur. Phys. J. B* (cond-mat/0307184).
- [86] R. Cohen and S. Havlin, *Phys. Rev. Lett.* **90**, 058701 (2003).
- [87] F. Chung and L. Lu, *Proc. Natl. Acad. Sci. USA* **99**, 15879 (2002).
- [88] S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin (cond-mat/0206467).
- [89] P. L. Krapivsky and S. Redner, *Phys. Rev. E* **63**, 066123 (2001).
- [90] D.-S. Lee, K.-I. Goh, B. Kahng and D. Kim (cond-mat/0312336).
- [91] T. E. Harris, *The Theory of Branching Processes* (Springer-Verlag, Berlin, 1963).
- [92] R. Otter, *Ann. Math. Statist.* **20**, 206 (1949).
- [93] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, *Phys. Rev. Lett.* **85**, 4625 (2000).
- [94] D. S. Callaway, M. E. J. Newman, S. H. Strogatz and D. J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).
- [95] M. Molloy and B. Reed, *Random Struct. Algorithms* **6**, 161 (1995).
- [96] K.-I. Goh, B. Kahng and D. Kim, *Phys. Rev. E* **64**, 051903 (2001).
- [97] I.J. Farkas, I. Derényi, A.-L. Barabási and T. Vicsek, *Phys. Rev. E* **64**, 026504 (2001).
- [98] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes and S. N. Samukhin, *Phys. Rev. E* **68**, 046109 (2003).
- [99] K. A. Eriksen, I. Simonsen, S. Maslov and K. Sneppen, *Phys. Rev. Lett.* **90**, 148701 (2003).
- [100] I. Simonsen, K. A. Eriksen, S. Maslov and K. Sneppen (cond-mat/0312476).
- [101] R. Albert, H. Jeong and A.-L. Barabási, *Nature* **406**, 378 (2000).
- [102] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, *Phys. Rev. Lett.* **86**, 3682 (2001).
- [103] B. A. Carreras, D. E. Newman, I. Dobson and A.B. Poole, *Hawaii International Conference on System Sciences, IEEE* (2001).
- [104] M. L. Sachtjen, B. A. Carreras and V. E. Lynch, *Phys. Rev. E* **61**, 4877 (2000).
- [105] D. J. Watts, *Proc. Nat. Acad. Sci. USA* **99**, 5766 (2002).
- [106] A. E. Motter and Y.-C. Lai, *Phys. Rev. E* **66**, 065102(R) (2002).
- [107] P. Holme and B. J. Kim, *Phys. Rev. E* **65**, 066109 (2002).
- [108] Y. Moreno, R. Pastor-Satorras, A. Vazquez and A. Vespignani, *Europhys. Letts.* **62**, 292 (2003).

- [109] P. Holme, B. J. Kim, C. N. Yoon and S. K. Han, Phys. Rev. E **65**, 056109 (2002).
- [110] K.-I. Goh, D.-S. Lee, B. Kahng and D. Kim, Phys. Rev. Lett. **91**, 148701 (2003).
- [111] F. Igloi and L. Turban, Phys. Rev. E **66**, 036140 (2002).
- [112] A. V. Goltsev, S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **67**, 026123 (2003).
- [113] S. N. Dorogovtsev, A. V. Goltsev and J. F. F. Mendes, (cond-mat/0310693).
- [114] L. Freeman, Sociometry **40**, 35 (1977).
- [115] M. E. J. Newman, Phys. Rev. E **64**, 016132 (2001).
- [116] U. Brandes, J. Math. Sociol. **25**, 163 (2001).
- [117] M. Barthélemy, Phys. Rev. Lett. **91**, 189803 (2003).
- [118] K.-I. Goh, C. M. Ghim, B. Kahng and D. Kim, Phys. Rev. Lett. **91**, 189804 (2003).
- [119] G. Bianconi and A. Capocci, Phys. Rev. Lett. **90**, 078701 (2003).
- [120] K.-I. Goh, E. Oh, B. Kahng and D. Kim, Proc. Natl. Acad. Sci. USA **99**, 12583 (2002).
- [121] K.-I. Goh, E. Oh, B. Kahng and D. Kim, Phys. Rev. E **67**, 017101 (2003).
- [122] R. Monasson, Eur. Phys. J. B **12**, 555 (1999).
- [123] J. D. Noh and H. Rieger (cond-mat/0307719).
- [124] M. E. J. Newman (cond-mat/0309045).
- [125] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002).
- [126] V. Gudkov, J. E. Johnson and S. Nussinov (cond-mat/0209111).
- [127] H. Zhou, Phys. Rev. E **67**, 061901 (2003).
- [128] F. Wu, B. A. Huberman (cond-mat/0310600).
- [129] J.-H. Kim, K.-I. Goh, B. Kahng and D. Kim, Phys. Rev. Lett. **91**, 058701 (2003).
- [130] L. A. Braunstein, S. V. Buldyrev, R. Cohen, S. Havlin and H.E. Stanley, Phys. Rev. Lett. **91**, (2003).
- [131] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
- [132] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001); Phys. Rev. E **63**, 066117 (2001); **65**, 035108 (2002).
- [133] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **65**, 036104 (2002).
- [134] R. Cohen, S. Havlin, D. ben-Avraham, Phys. Rev. Lett. **91**, 247901 (2003).
- [135] Z. Dezsó and A.-L. Barabási, Phys. Rev. E **65**, 055103(R) (2002).
- [136] T. Nishikawa, A. E. Motter, Y.-C. Lai and F. C. Hoppensteadt, Phys. Rev. Lett. **91**, 014101 (2003).
- [137] H. Hong, M. Y. Choi and B. J. Kim, Phys. Rev. E **65**, 026139 (2002).
- [138] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **80**, 2109 (1998); K. S. Fink, G. Johnson, T. Carroll, D. Mar and L. Pecora, Phys. Rev. E **61**, 5080 (2000); M. Barahona and L. M. Pecora, Phys. Rev. Lett. **89**, 054101 (2002).
- [139] S. Jalan and R.E. Amritkar (nlin.AO/0201051).
- [140] J. M. Kleinberg and S. Lawrence, Science **294**, 1849 (2001).
- [141] J. M. Kleinberg, J. ACM **46**, 604 (1999).
- [142] J. M. Kleinberg, Nature **406**, 845 (2000).
- [143] B.J. Kim, C.N. Yoon, S.K. Han and H. Jeong, Phys. Rev. E **65**, 027103 (2001).
- [144] J. E. S. Socolar and S. A. Kauffman, Phys. Rev. Lett. **90**, 068702 (2003).
- [145] D. Bray, Science **301**, 1866 (2003).
- [146] A.-L. Barabási and Z. N. Oltvai, Nature Genetics **5**, Feb (2004).
- [147] J.-P. Onnela, K. Kaski and J. Kertész (cond-mat/0312682).
- [148] H.-J. Kim, Y. Lee, I-m Kim and B. Kahng, J. Kor. Phys. Soc. **40**, 1105 (2002); J. Phys. Soc. Japan **71**, 2133 (2002).
- [149] G. Bonanno, G. Caldarelli, F. Lillo, S. Micciché, N. Vandewalle, R. N. Mantegna (cond-mat/0401300).
- [150] D. Stauffer, A. Aharony, L. da Fontoura Costa and J. Adler (cond-mat/0212601).

Complex Networks: Structure and Dynamics

B. KAHNG,* K.-I. GOH, D.-S. LEE and D. KIM

School of Physics, Seoul National University NS50, Seoul 151-747

(Received 27 January 2004)

Recently, complex network has drawn considerable attentions as a basic tool to help us to understand a variety of phenomena in interdisciplinary fields such as biology, computer science, economics, mathematics, physics, and sociology. One of the empirical facts commonly appeared in empirical data of diverse fields is the heterogeneity in the connectivity distribution of each node, following a power-law distribution. This behavior is attractive as a new emerging behavior in complex systems. Here we review developments during last 3-4 years in this field, including structural properties of real-world networks and artificial models, the evolution of network topologies, dynamic problems on such complex networks.

PACS numbers: 64.60.-i

Keywords: Complex systems, Network, Power-law

*E-mail: kahng@phya.snu.ac.kr