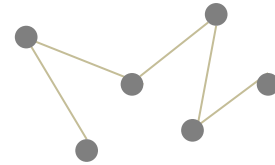


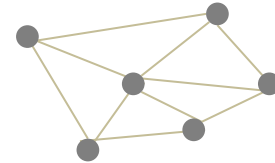
Park, Jung Eun @ ComplexitySpin 151128



어떠한 구성요소(부분)로 이루어졌는지 뿐 아니라,



어떤 방식으로 구성되었는지 역시 중요하다.

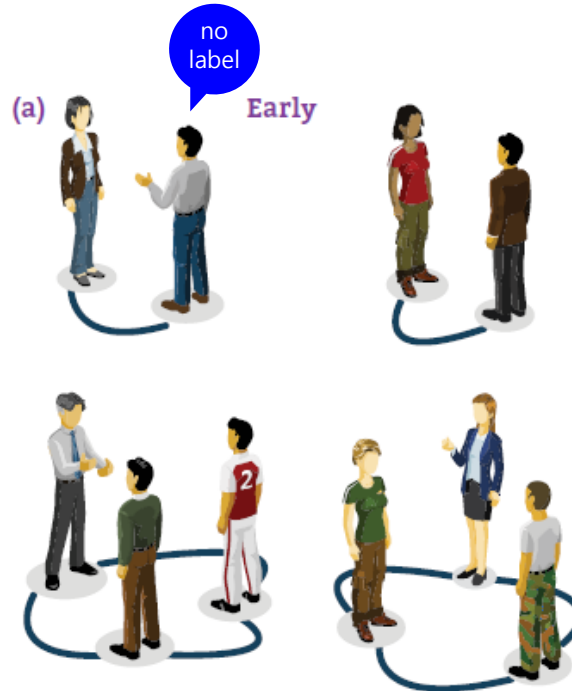


에르되스-레니 모델(Erdős-Rényi model) : Random Network

서로 아는 사람이 없는 100명을 파티에 초대하고, 와인과 치즈를 주면 이들은 서로 안면을 트기 시작한다.

손님 중 한 명에게
'라벨이 붙지 않은 와인이 빨간 라벨이 붙은 와인보다 훨씬 좋은 최고급 와인이다'라고 이야기하고,
새로 사권 손님들과만 이 사실을 공유하라고 한다.

파티가 끝나기 전까지 최고급 와인은 무사할 수 있을까?



에르되스와 레니 - 랜덤 네트워크

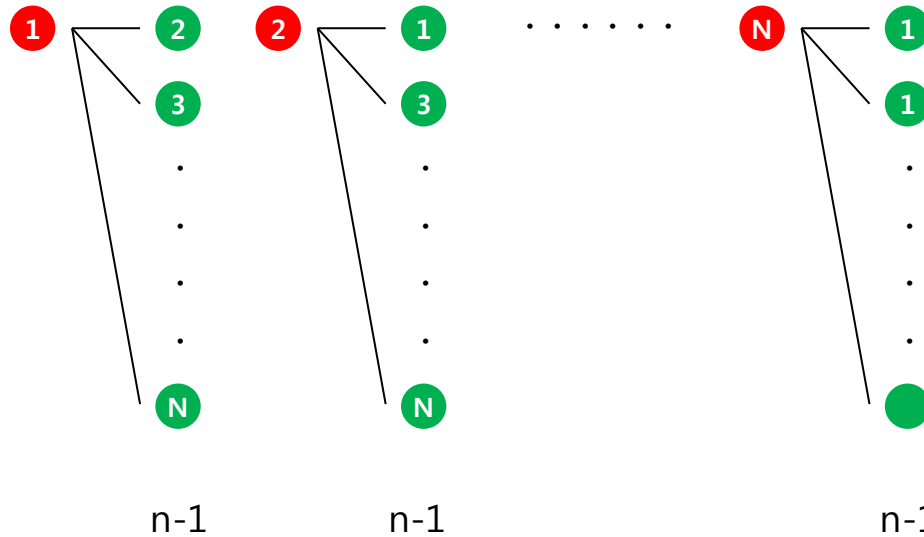
- E.D. Gilbert가 같은 시기에 독립적으로 같은 내용의 연구를 발표했으나 E-R이 워낙 빵 뚫기에 그들을 founder로 본다.
- 모든 노드는 동일한 조건을 가지며, 고정되어 있다(성장 x)
- 링크는 무작위적으로 발생한다. (node i 와 j 사이에 link가 발생할 확률은 전체 네트워크에서 동일)
- 처음에는 단절된 소그룹들만이 존재하지만, 각 노드가 평균적으로 하나의 링크를 가질 정도로 충분한 수의 링크를 추가하다 보면(임계치), 소그룹 사이를 연결하는 커다란 클러스터가 발생한다.
- 결과적으로 모든 노드는 연결된다.
→ 랜덤 네트워크에서 전체가 모두 연결되기 위해서는 노드당 오직 하나의 링크만 있어도 충분하다.



랜덤 네트워크 만드는 방법

1. N개의 고립된(서로 전혀 연결되지 않은) 노드를 만든다.
2. 두 노드 사이에 링크가 생길 확률 p를 설정한다.
3. 임의로 두개의 노드를 선택한다.
4. 0~1사이에서 랜덤하게 수를 뽑는다.
5. 이 수가 p보다 크면 링크를 만들고, 작으면 다시 3부터 반복한다.

N개의 노드에서 임의의 두 개 노드를 선택하는 경우의 수 = $\frac{N(N-1)}{2}$



$n \times (n-1)$

1-2, 2-1로 중복되므로
undirected network에서는 2로 나눔

이는 N개의 공이 든 상자에서 순서에 상관없이 두 개의 공을 선택하는 경우의 수와 같다.

$$\rightarrow {}_N C_2 = \frac{N!}{2!(N-2)!} = \frac{N(N-1)(N-2)!}{2(N-2)!} = \frac{N(N-1)}{2}$$

※ 베르누이 시행

- 베르누이 시행(Bernoulli trials) : 어떤 실험 또는 시행의 결과가 두 가지 가운데 하나만 나오는 경우 (성공/실패, 있음/없음, 동전 앞/뒤 등), 이런 시행을 n회 반복할 경우에 매회 확률이 동일하고 각 시행이 독립적이면 베르누이 시행이라고 한다.

- 이 경우 한 쪽의 발생확률을 p라고 하면 다른 한 쪽의 확률은 1-p이며, 확률질량함수(probability mass function)는 다음과 같이 쓸 수 있다.

$$p(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$$

- 이 때 X는 성공률 p인 베르누이 분포(Bernoulli distribution)를 따른다고 하며, 아래와 같이 나타낸다.

$$X \sim b(p), \quad 0 \leq p \leq 1$$

- X의 기대값(expected value)과 분산(variance) 다음과 같다

$$\mu = E(X) = \sum_{x=0}^1 xp^x(1-p)^{1-x} = (0)(1-p) + (1)(p) = p$$

$$\begin{aligned} \sigma^2 = var(X) &= \sum_{x=0}^1 (x-p)^2 p^x(1-p)^{1-x} \\ &= p^2(1-p) + (1-p)^2 p = p(1-p) \end{aligned}$$

※ 이항분포(binomial distribution)

- 베르누이 분포는 $n = 1$ 인 경우를 말하며, 이를 독립적으로 n 회 실시한 경우의 분포를 이항분포라 한다.
- 만약 확률변수 X 를 n 회 베르누이 시행에서 관측된 성공횟수로 정의한다면, X 의 가능한 값들은 $0, 1, 2, \dots, n$ 일 것이다. 만약 $x=0, 1, 2, \dots, n$ 일 때 x 번의 성공이 일어났다면, $n-x$ 번의 실패가 일어난 것이다. 따라서 n 번의 시행에서 x 번의 성공에 대해 X 가 x 값을 취할 경우의 수는 다음과 같다.

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

- 이 시행들은 독립이고 각각의 시행에서 성공과 실패의 확률은 p 와 $1-p$ 이고, 각 경우에 대한 확률은 $p^x(1-p)^{n-x}$ 이므로,

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n$$

이항분포는 베르누이 시행을 독립적으로 n 번 시행한 결과이므로
 $X \sim B(n, p)$ 인 X 에 대해서

$$X = k + k + \dots + k = nk$$

곧, 기대값과 분산의 성질에 따라, $X \sim B(n, p)$ 인 확률변수 X 에 대해서 다음이 성립합니다. (독립인 A, B 사건에 대해 $\text{Var}(A+B) = \text{Var}(A) + \text{Var}(B)$ 성립)

$$E(X) = E(nk) = nE(k) = np$$

$$\text{Var}(X) = \text{Var}(k_1) + \dots + \text{Var}(k_n) = n \text{Var}(k) = np(1-p)$$

3.3 Number of Links : 랜덤 네트워크가 L개의 링크를 갖게 될 확률

1. 네트워크에 생길 수 있는 $\frac{N(N-1)}{2}$ 개의 링크 중 L개가 존재할(연결될) 확률 = p^L
2. 나머지 $\frac{N(N-1)}{2} - L$ 개의 링크는 존재하지 않을 확률 = $(1-p)^{\frac{N(N-1)}{2} - L}$
3. $\frac{N(N-1)}{2}$ 개의 링크 중 L개를 선택할 수 있는 경우의 수 = $\binom{\frac{N(N-1)}{2}}{L}$

$$\times \binom{N}{r} = {}_N C_r = \frac{N!}{r!(N-r)!}$$

4. 따라서, 어떤 랜덤 네트워크가 L개의 링크를 갖게 될 확률 = $\binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2} - L}$

5. 위의 식은 이항분포(binomial distribution)를 따르므로, 어떤 랜덤 네트워크가 갖는 링크의 개수 L의 기대값은

$$\langle L \rangle = \sum_{L=0}^{\frac{N(N-1)}{2}} L p_L = p \frac{N(N-1)}{2}$$

※ 이항분포의 기대값은 np

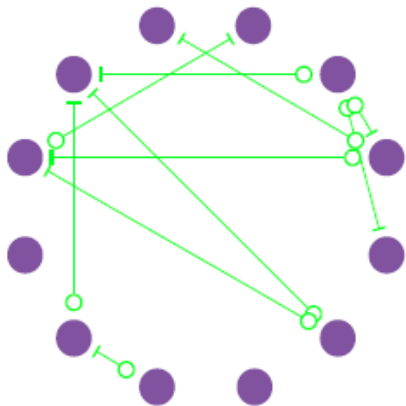
6. 이 네트워크의 평균 degree k의 기대값은 $\langle k \rangle = \frac{2\langle L \rangle}{N} = \frac{2}{N} \times p \frac{N(N-1)}{2} = p(N-1)$

7. N개의 노드를 가지는 랜덤 네트워크에서 임의의 두 노드가 연결될 확률 p를 $0 \rightarrow 1$ 로 점진적으로 증가시키면, $\langle L \rangle$ 은 $0 \sim L_{\max}$ ($= \frac{N(N-1)}{2}$)로, $\langle k \rangle$ 는 $0 \sim N-1$ 로 linear하게 변화한다.

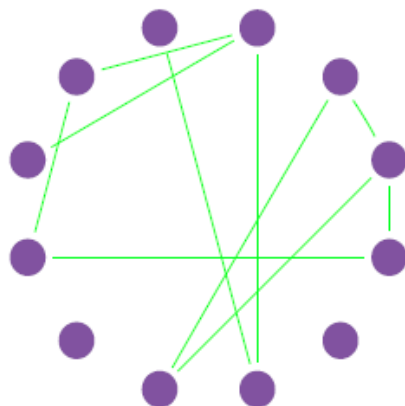
랜덤 네트워크의 N, p 가 동일하더라도 L , isolated node, 평균 degree 등이 다르게 나타날 수 있다.

$p = 1/6, N = 12$

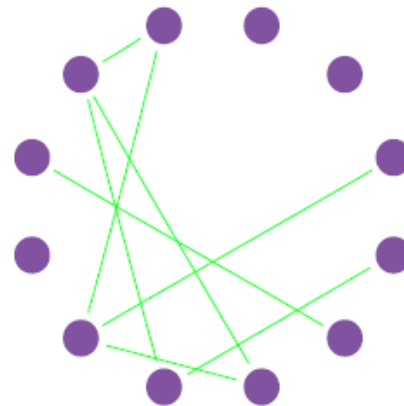
$L = 10$



$L = 10$

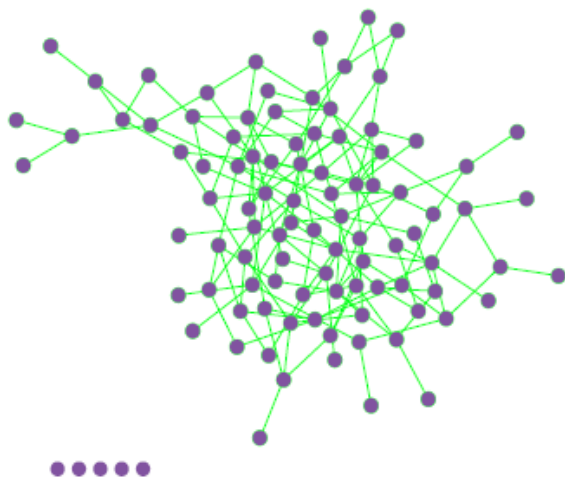


$L = 8$

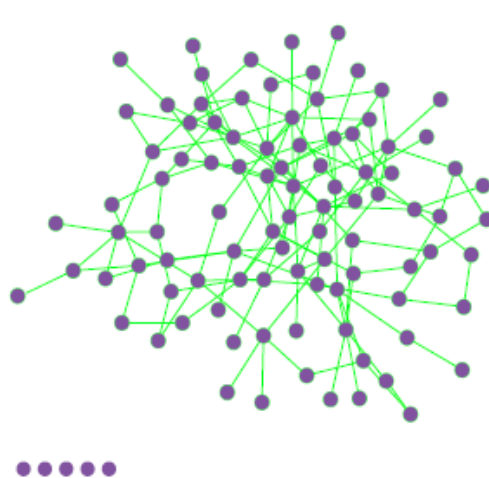


$p = 0.03, N = 100$

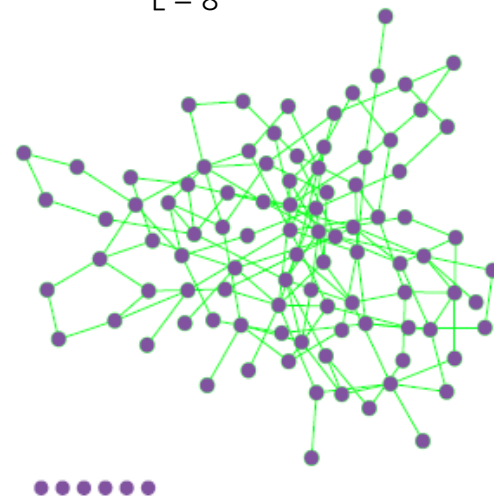
$L = 10$



$L = 10$



$L = 8$



3.4 Degree Distribution: 랜덤 네트워크에서 각 노드의 degree 분포(Bollobás, 1981)

N개의 노드를 가지는 랜덤 네트워크에서 각 노드가 갖는 degree 의 분포는 어떻게 나타날까?

임의의 노드 i 가 가질 수 있는 degree의 최대값인 N-1 중에,

1. k개의 링크가 존재할 (degree = k) 확률 = p^k
2. 나머지 (N - 1 - k) 개의 링크는 존재하지 않을 확률 = $(1-p)^{N-1-k}$
3. 최대 N-1개의 링크 중 k개를 선택할 수 있는 경우의 수 = $\binom{N-1}{k} = {}_{N-1}C_k$

4. 따라서, 랜덤 네트워크의 degree distribution은 다음과 같은 이항분포를 따른다.

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

이 분포의 형태는 N과 p에 의해 결정되며, 각각의 경우에 k 와 σ 를 계산할 수 있다.

포아송 분포(Poisson Distribution)

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	$\langle k \rangle$
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.34
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

Table 2.1
Canonical Network Maps

대부분의 real network는 성긴(sparse) 구조이다. 즉, $\langle k \rangle \ll N$ 이다.
그런데 $\langle k \rangle \ll N$ 인 경우 앞의 이항분포를 포아송 분포로 근사시키는 것이 가능하다.

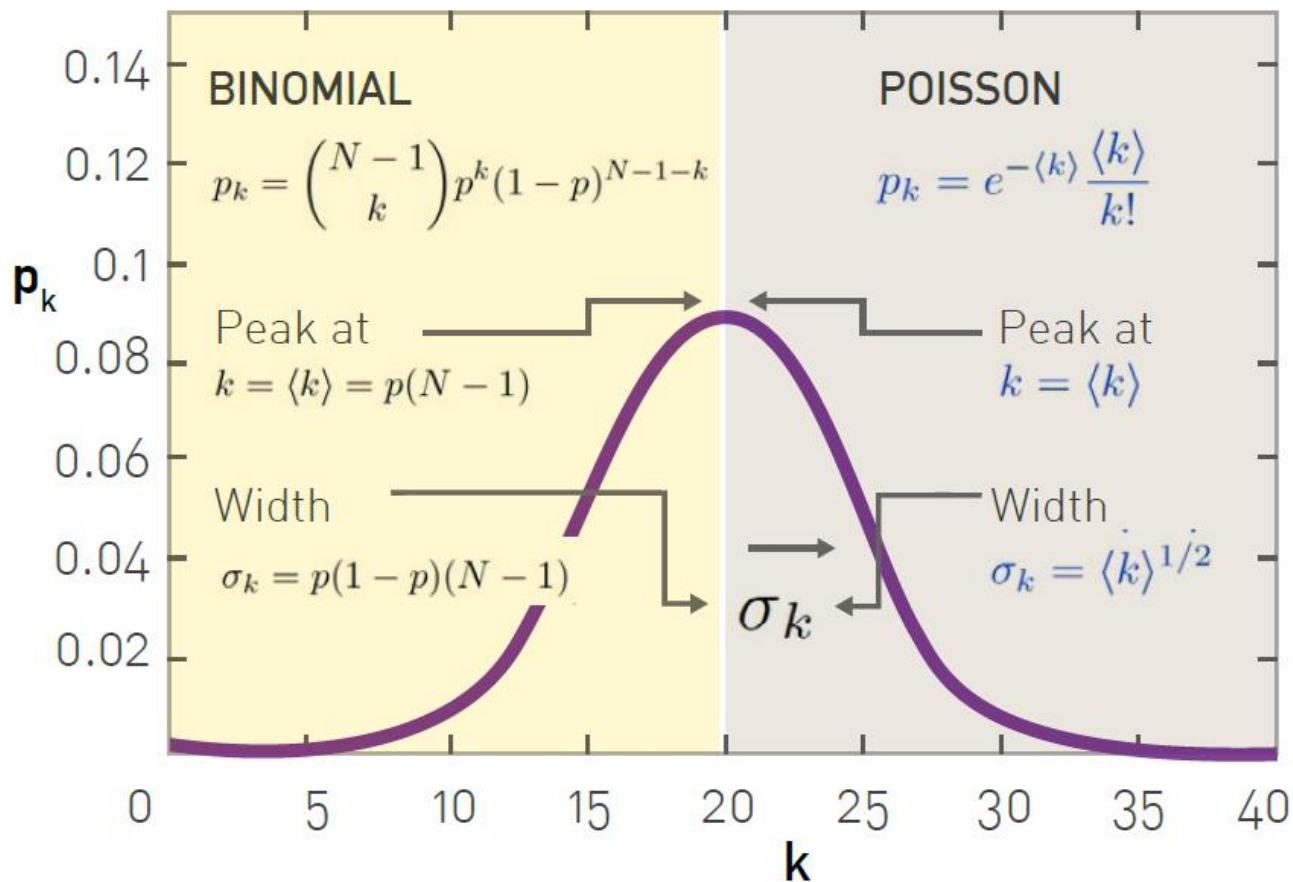
$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \longrightarrow p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

(그래? 그럼 real network는 포아송 분포를 따를지도...? 일단 넘어가자)

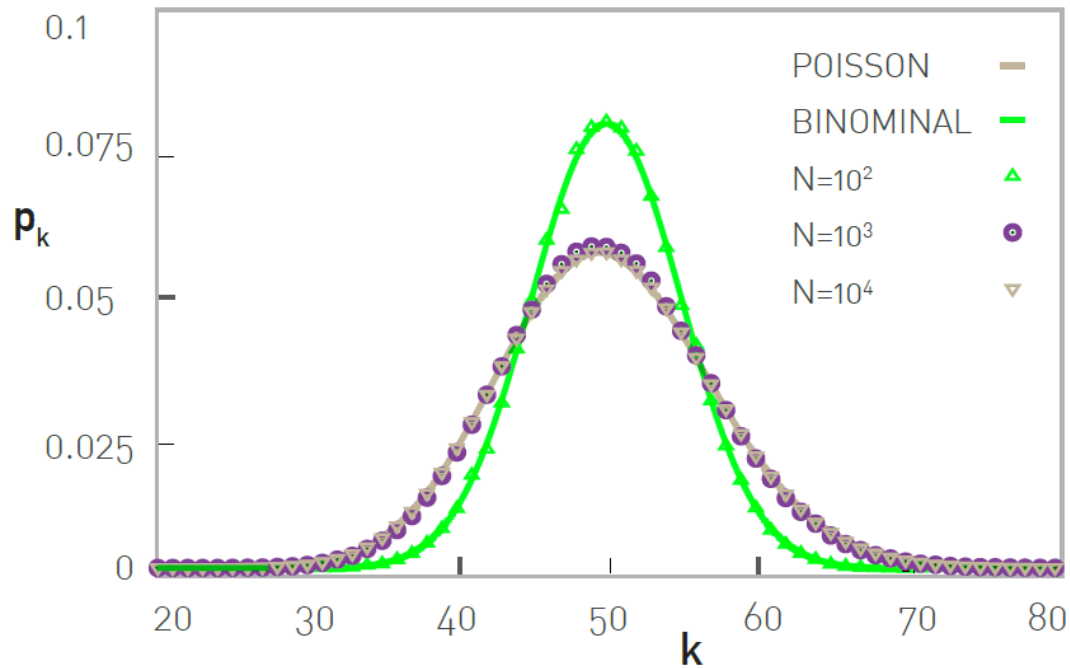
이항분포 vs. 포아송 분포

포아송 분포가 이항분포에 비해 더 유리한 점이 무엇일까?

(힌트 : 실용적인 측면, 즉 어딘가에 써먹을 경우를 생각해 보자)



네트워크 규모(N)에 따른 랜덤 네트워크의 degree distribution



small network (binomial) : k 의 평균, 분산, 링크 확률 등의 척도들이 N 과 p 의 영향을 받는다.

large network (poisson) : degree의 편차가 이항분포보다 크다. 그리고, 척도들을 k 하나로 나타낼 수 있다.

일단 랜덤 네트워크의 degree distribution은 포아송 분포를 따른다고 치고 넘어가자(바라바시 아저씨가 그랬음)

3.5 Real Networks are NOT Poisson

A ㅏ

(그래 그렇게 쉽게 풀릴리가 있나..)

한편 여기서부터 나의 수학력은 한계를 맞이했습니다.

지수함수와 원주율과 자연로그가 나오면서부터 대략 유체가 이탈되고...

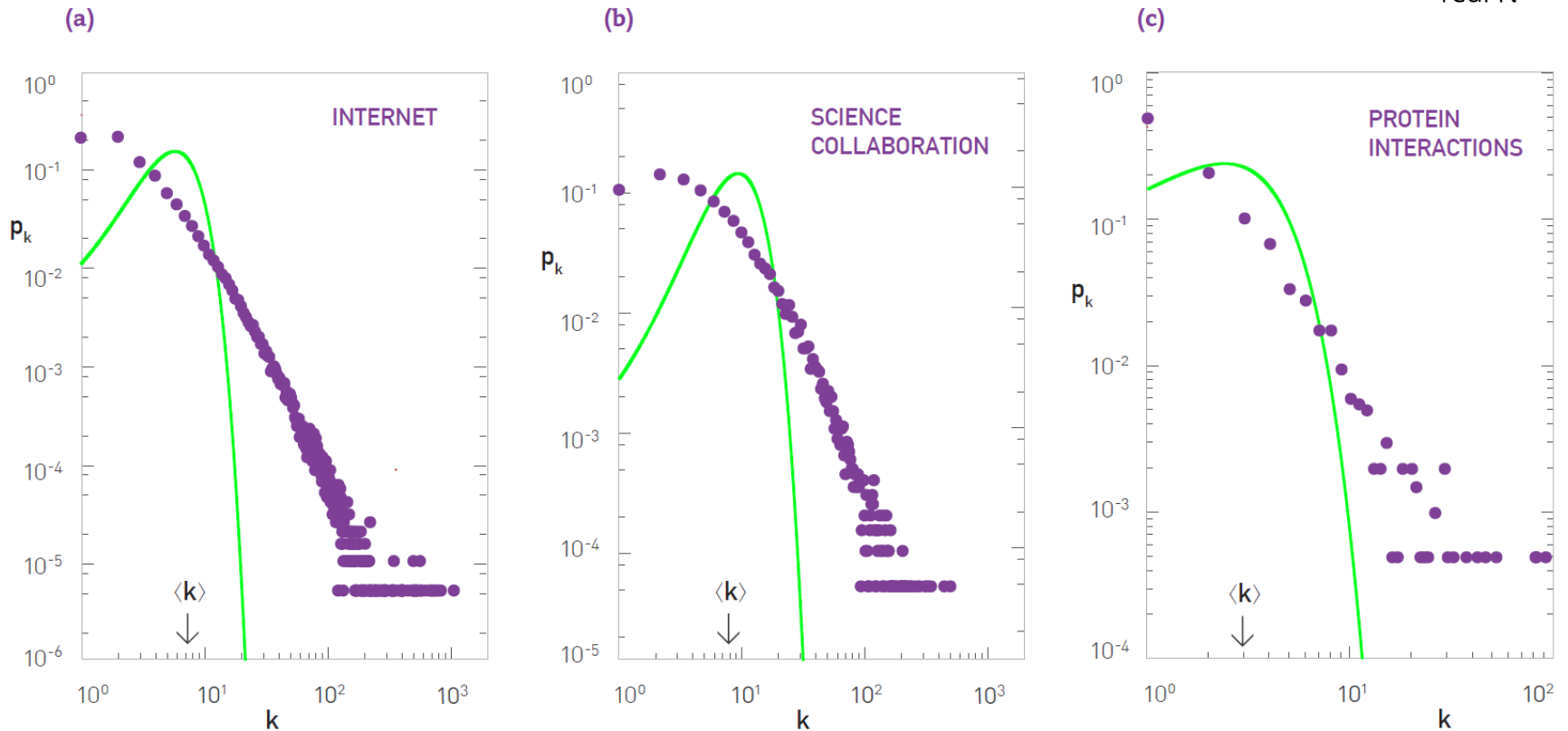
여기서부터는 개념만 설명하는 것으로 합시다.

random network vs. real network

random network로는 실제 네트워크의 현상을 설명하지 못한다..
랜덤 N에서는 대부분의 노드가 비슷한 수의 링크를 가지지만,
실제 N에서는 랜덤 N에서보다 더 많은 노드가 outlier를 가지며, k 값의 범위도 크다

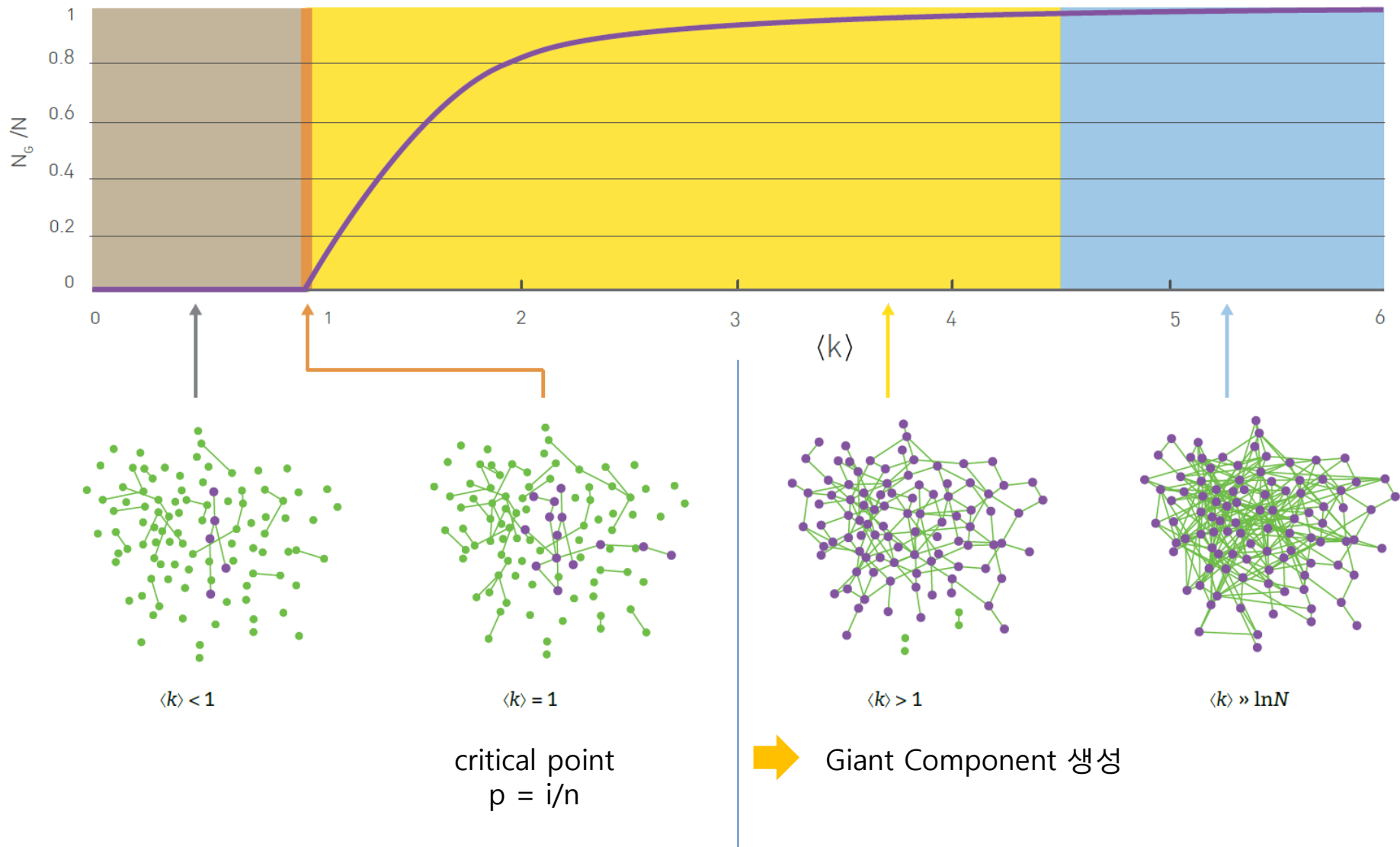
사실 에르되스와 무작위 모델을 제시한 것은 일관성과 수학적 아름다움 때문이며
실제와는 다를 것이라는 것을 그들도 알고 있었다고 한다.

— random N
- - - real N



3.7 Evolution of a random network

처음에 얘기한 와인파티에서,
“랜덤 네트워크에서 전체가 모두 연결되기 위해서는 노드당 오직 하나의 링크만 있어도 충분하다.” 고 했다.



한편 여기서부터 나의 수학력은 한계를 맞이했습니다.

지수함수와 원주율과 자연로그가 나오면서부터 대략 유체가 이탈되고...

여기서부터는 개념만 설명하는 것으로 합시다.

여섯 단계의 분리 : Milgram의 편지 실험

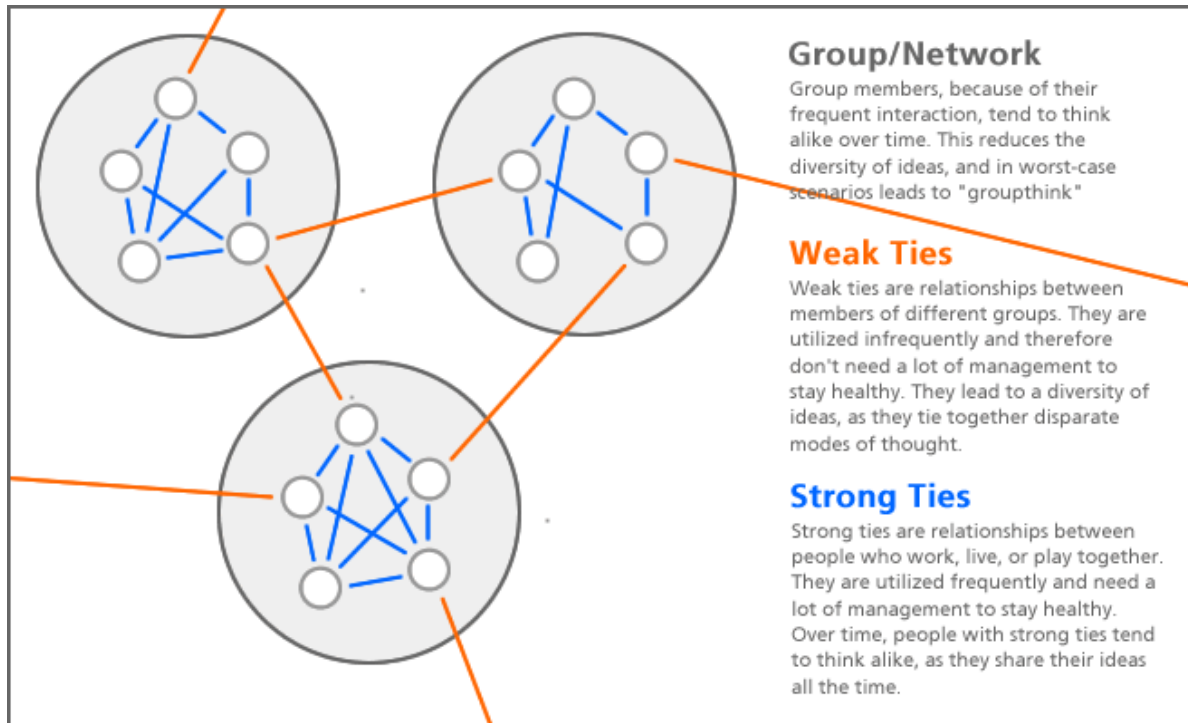
미국에 사는 임의의 두 사람을 연결하기 위해서 그들 사이에 얼마나 많은 지인이 필요한가?

- 위치타와 오마하 주민을 무작위로 선정, 한 주민에게서 시작해 목표인물에게 편지를 전달하는 실험
- 목표인물을 알고 있다면 직접 전달
- 모른다면 그 인물을 알만한 사람에게 우편 송부
- 160개 중 42개의 편지가 목표인물에게 도달함
- 거친 사람은 2명~12명, median = 5.5명 → Six Degrees of Separation
- 60억 노드로 이루어진 네트워크에서 임의의 노드 한쌍을 선택했을 때 그들 간의 거리는 평균적으로 6단계밖에 되지 않음
→ 좁은 세상(Small World) 네트워크

그라노베터 - 약한 연결의 힘(The Strength of Weak Ties)

약한 연결이 강한 친분관계보다 중요

- 관리직/전문직 종사자 수십여 명에게 누가 현재의 직업을 찾는 데 도움을 주었는가를 물었을 때,
- 대부분이 친한 친구가 아니라 ‘그냥 아는 사람(acquaintances)’이라고 답함
- 정보의 전파에 있어서 약한 사회적 연결이 강한 사회적 연결보다 중요하다.



와츠-스트로가츠의 좁은 세상 네트워크(Watts & Strogatz : Small World network)

- Real-world network : high level clustering, small distances between most nodes
- “나의 친구 두 사람이 서로 알 확률은 얼마일까?” → random network에서는 모든 node에서 동일함
- 그라노베퍼 : 고도로 조밀한 클러스터 + 소수의 약한 연결

$$0 \quad \leftarrow \quad \text{Clustering Coefficient} = \frac{\text{실제 내 친구들 사이에 나타나는 link 개수}}{\text{내 친구들도 모두 서로 친구일 때 link 개수}} \quad \leftarrow \quad 1$$

클러스터링 계수, 결집계수

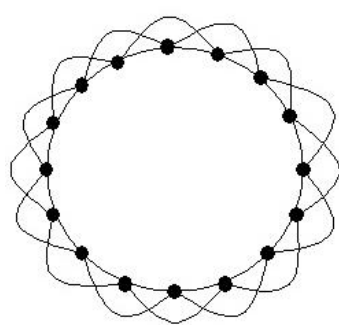
나만
다중생활

→ 내 친구들의 집단이 얼마나 조밀하게 짜여져 있는가?

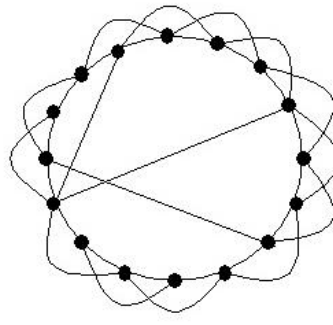
서로
다 안다

Watts - Strogatz : Small World network

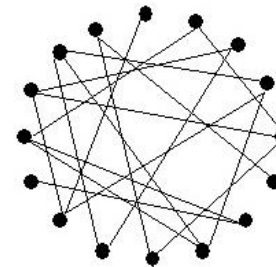
- N 개의 노드가 배열된 lattice 구조의 네트워크에서, 각 노드들은 최인접노드와 그 다음 최인접 노드와 연결(neighbor)
 - P의 확률로 link를 선택, 양 노드 중 하나를 선택하여 고정하고 다른 한쪽은 네트워크 전체 노드 중 하나를 선택하여 rewiring
 - 이미 연결된 노드는 배제
- 링크를 몇 개만 추가해도 네트워크의 조밀도(클러스터링 계수)가 크게 변하지 않으면서 노드간의 평균 거리는 급격하게 줄어든다.



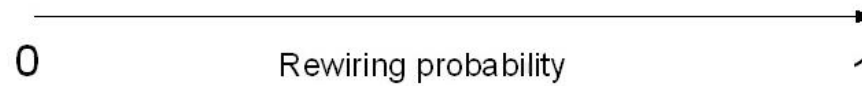
ordered



small-world



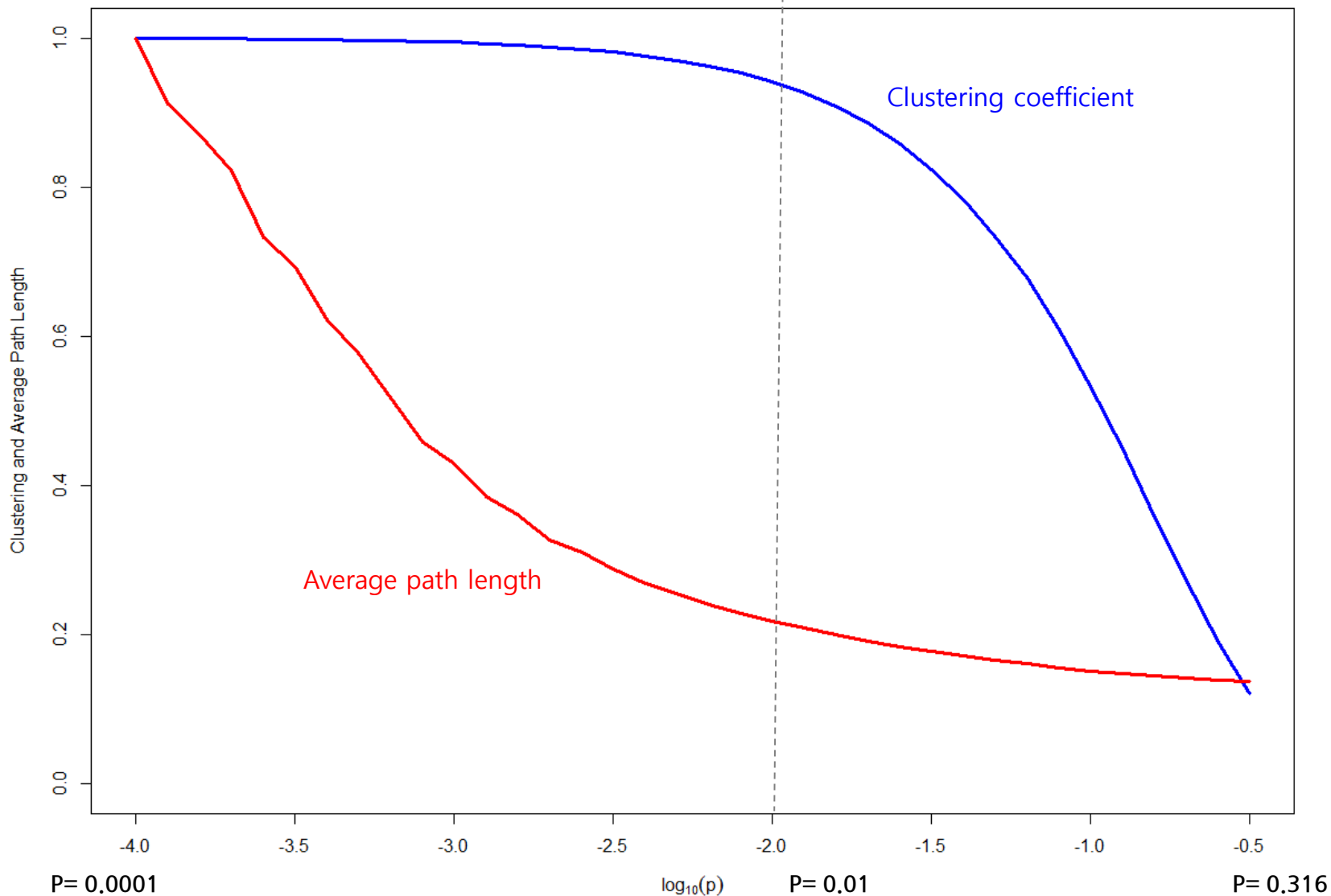
random



N = 1000, r = 10인 small world network를 100번 생성 → 클러스터링 계수와 평균 path 길이 변화의 양상을 확인

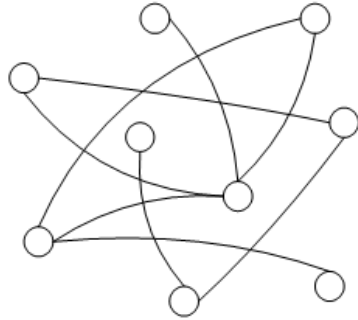
Link 생성 확률 p가 매우 작을 때
큰 클러스터링 계수를 유지하면서
소수의 링크만으로 평균 최단경로길이가 급격하게 감소

링크수와 링크생성확률이 증가함에 따라 더 이상 최
단경로길이가 짧아지지 않는 임계치를 넘을 때까지 높
은 클러스터링 계수 유지



Random Network vs. Small World Network

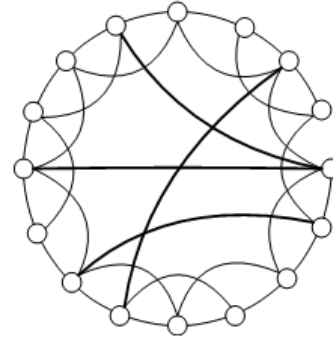
무작위 네트워크



Erdős-Rényi

- 전체가 모두 연결되기 위해서는 노드당 오직 하나의 링크만 있어도 충분하다
- 네트워크가 커지면 무작위로 링크를 부여하더라도 각 노드가 거의 같은 수의 링크를 가지게 된다
- 모든 노드는 처음부터 주어져 고정되며
- 네트워크의 생애동안 불변
- 모든 노드는 동등하다

좁은 세상 네트워크



Watts - Strogatz

- 링크를 몇 개만 추가해도 노드 간의 평균거리가 급격하게 줄어든다.
- 이 때 클러스터링 계수는 크게 변화가 없음
- 모든 노드는 동등하지 않다
- 링크의 재배열(Rewiring)

무작위적 세계관의 폐기

- 말콤 글래드웰, *티핑 포인트*, 21세기북스, 2004
- ‘내가 얼마나 사교적인(social) 사람인가?’를 알아볼 수 있는 테스트(총 400명):
- 맨해튼 전화번호부에서 뽑은 248개의 성(性) 목록을 주고 그러한 성을 가진 사람을 몇 명이나 알고 있는지 점수화
- 성씨는 중복 가능

sample	mean	min	max
맨해튼 시립대학 학생들	21	2	95
고학력 백인교수 무작위 표본	39	9	118
비슷한 연령대, 교육수준, 소득수준 표본		16	108

애네 뭐야 무서워..

“모든 계층을 막론하고 친구나 아는 사람을 만드는 데 있어서 극히 예외적인 솜씨를 가진 소수의 사람들이 있다. 그들은 Connector라 할 수 있다.”

무작위적 세계관의 폐기

1. Web에서의 Connector : Hub

- 웹에서는 소수의 허브가 링크의 대부분을 crawling
- 허브는 전체 네트워크 구조를 지배하며, 그것을 좁은 세상으로 만드는 역할을 한다

2. Pareto(80/20)의 법칙 : 소수의 큰 큰 사건이 현상의 대부분을 포괄한다.

ex. 전체 돈의 80%는 전체 인구의 20%가 소유한다

→ 이와 같은 현상은 전통적인 무작위 네트워크에서는 설명할 수 없음

그래서 scale-free network로 뽐!