Collective dynamics of “small-world” networks

Duncan J. Watts & Steven H. Strogatz

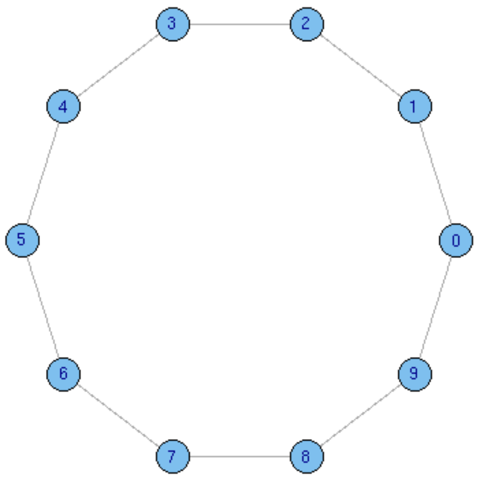
Regular Networks (출처) ; http://geza.kzoo.edu/~csardi/module/html/regular.html

1. Rings

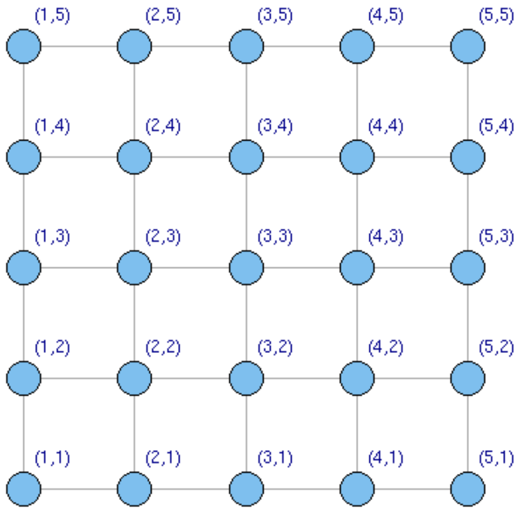
A ring is a connected graph in which each vertex is connected to exactly two other vertices.

2. Lattices

A lattice is a graph in which the vertices are placed on a grid and the neighboring vertices are connected by an edge. A one dimensional lattice is like a ring, only it is not circular, the circle is not closed. A two dimensional lattice can be seen in the following picture:



Ring



Lattice

3. Trees

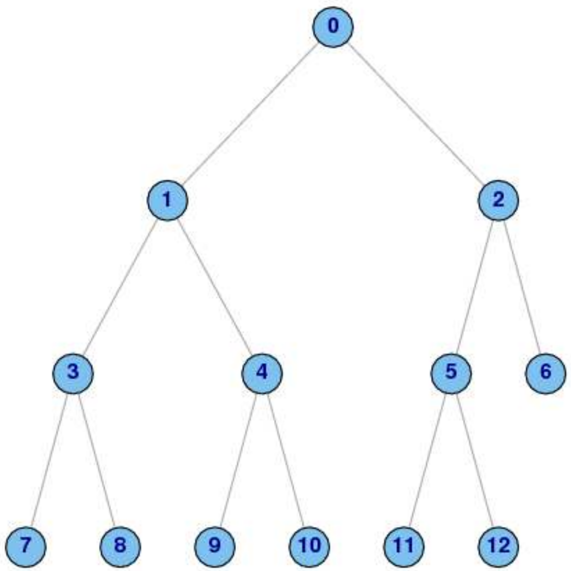
A tree is a connected graph which contains no circles (cycles). A tree graph is usually plotted "tree-like" with its root on the top and then its branches going downward. (Hence its name.) The top vertex is called the "root" and the vertices at the next lower level are called the children of the root. In general the neighbors of a vertex at a lower level are called the children of that vertex.

4. Stars

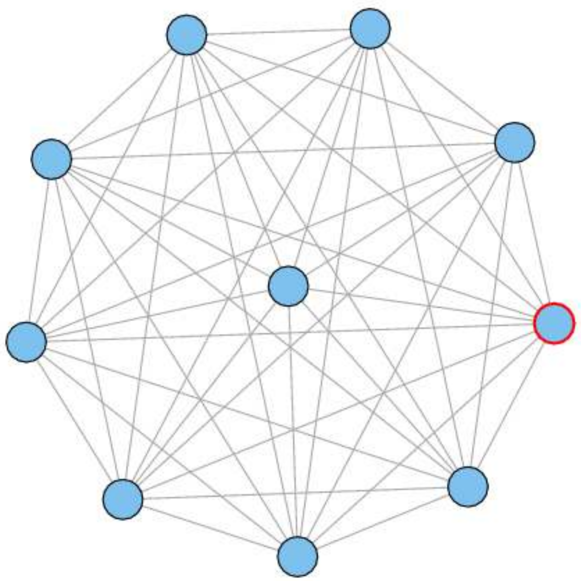
A star graph is a special tree, where every vertex is connected to the root.

5. Full graph

In a full graph every possible edge is realized, ie. there is an edge between every pair of vertices.



Tree



Full Graph

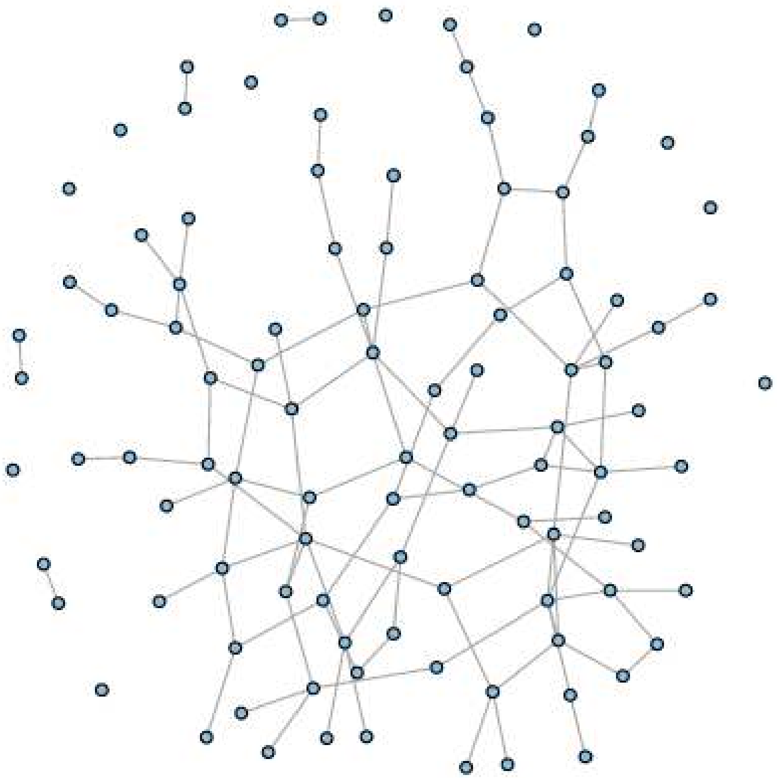
Edge 개수:

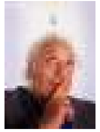
$v \cdot (v-1)/2$

v ; vertices 개수

Erdõs-Rényi random graphs

G(n,p) graphs are generated this way: the graph contains n vertices. Then for every pair of vertices with probability p an edge is drawn connecting them. Below is a G(n,p) graph with n=100 and p=2/100.





Regular 그래프, Random 그래프, 상식적으로 생각해 보자 ...

Regular 그래프를 "규칙적인 그래프", Random 그래프를 제멋대로인 "무작위 그래프"라 말해보자. 무엇이 규칙적이라는 것이지? Vertices 모양이 규칙적이라는 것인가? Vertices 숫자가? Vertices 내용이? 그래프에서 vertices 모양으로 그래프의 특성을 얘기하는가? 별로... Vertices의 개수가 많고, 적음에 따라 규칙적, 무작위라고 말할 수 있을까? 별로... Vertices의 내용으로? 아마도 그럴 수 있을지도 모르겠다. 그렇지만 이렇게 vertices들이 지니고 있는 content의 내용으로 그래프를 규칙적, 무작위라고 얘기하려면 content를 파악하는 것이 먼저 선결되어야 하기에 이는 퍽 어려운 얘기다. 따라서 아직은 vertices의 특성을 갖고 "규칙적이다, 제멋대로 무작위이다" 이렇게 따지기가 어렵다.

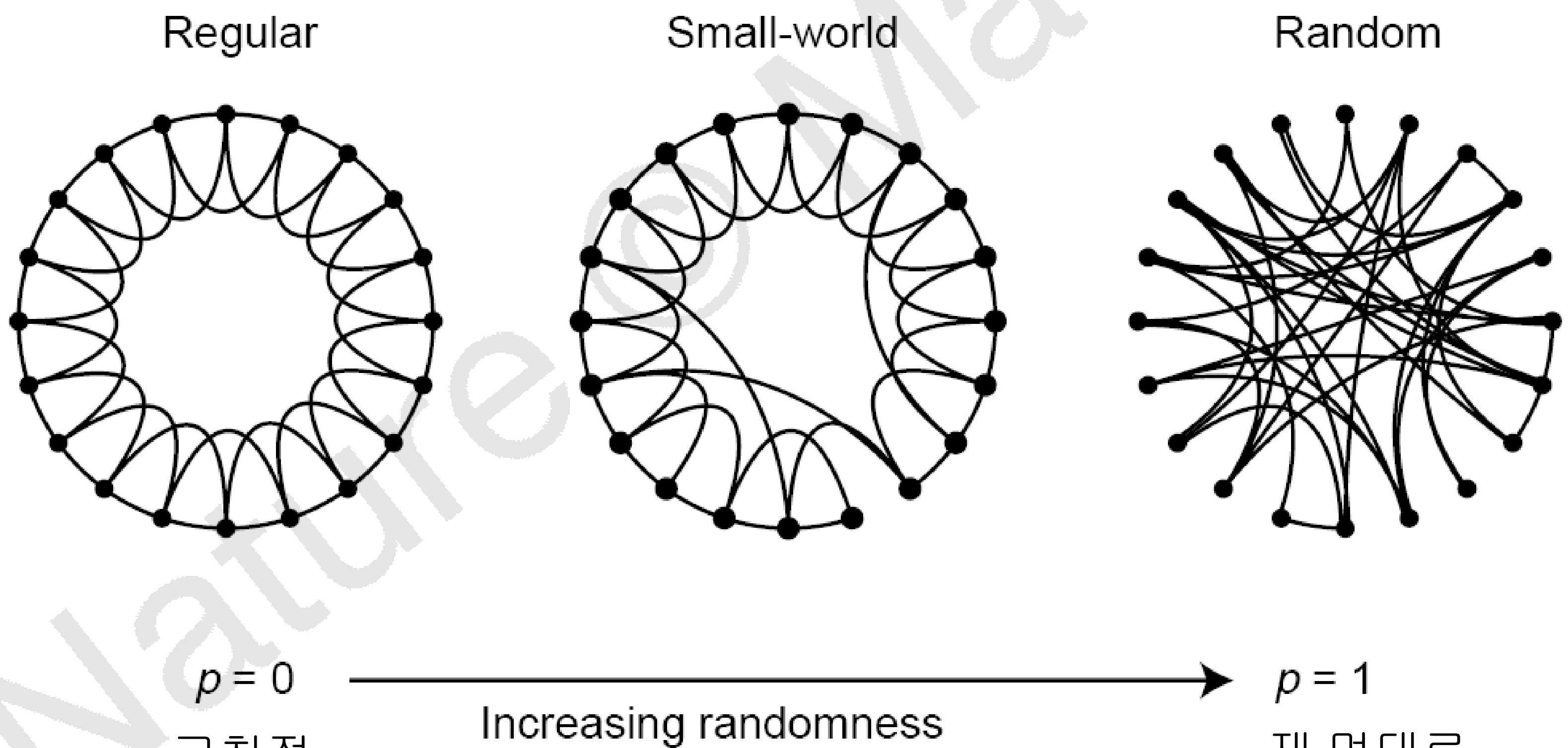
그러니, 그래프가 "규칙적이다" 또는 "무작위이다" 하는 것은 vertices를 연결하는 edge들의 패턴을 보고 하는 말이다. 좋다... 그러면 edge 들의 연결 패턴이 어떠하면 규칙적이라는 말인가? 정답은, 여러분이 반복되는 규칙적인 패턴을 발견할 수 있으면 규칙적인 그래프이다.

그러면 반복적으로 되풀이되는 규칙적인 edge들의 연결 패턴은 보통 어떤 특성이 있어야 보일까? 멀리 떨어진 vertices 들을 연결하는 edge들을 보고 쉽게 패턴을 발견하기 쉬운가? 아니다. 패턴들은 가까이 있는 vertices들을 연결하는 edge 들에서 쉽게 발견할 수 있다. 패턴이라는 말 자체가 어떤 지역적인 특성이 반복해서 전체적으로 나타난다는 말이다.

즉, Regular 그래프란 서로 가까이 위치한 vertices 들을 연결하는 edge들의 모양(structure, topology)이 전체 그래프에 걸쳐 계속, 반복적으로 나타나는 형태의 그래프라 말할 수 있다.

small-world" networks 논문의 주제

네트워크가 아래의 Regular 형태와 Random 형태의 중간의 형태 (small world)를 갖을 때, characteristic path length L(p) 와 clustering coefficient C(p) 개념을 도입해 네트워크가 regular에서 random 으로 나아갈 때 L과 C가 어떻게 변화하는 가를 살핀다. C가 높아 끼리끼리 노는 경향이 강한 네트워크에서 약간의 원거리 링크가 생겨도 금방 node간의 거리 (L)이 작아진다는 것을 밝힌다.



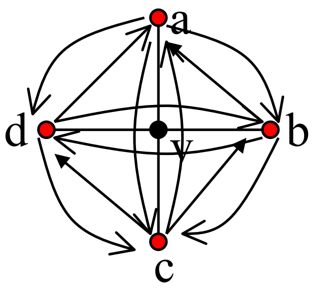
규칙적

제 멋대로,
무작위

네트워크의 “vertices간 평균 거리”와 “Clustering 정도” 측정 지표

L : Characteristic path length. Network내의 임의의 두 vertices 간 일반적 거리. 즉, 임의의 두 vertex를 선정했을 때 이들간의 shortest path length. 앞서 “graph structure in the web” 에서의 average connected distance와 같은 것임. L(p)는 p가 변화함에 따른 L의 값을 나타낸다.

C : Clustering Coefficient. 네트워크의 vertices들이 얼마나 끼리끼리 노는가, 즉 군집성을 나타낸다. 어떤 vertex인 v가 Kv 개의 neighbor (v와 edge로 직접 연결된 다른 vertices) 이 있다고 하자. 이 때, v의 neighbor를 v의 친구라 하고, v의 친구들이 서로 얼마나 직접 친구 (edge로 직접 연결) 되었나를 보아 이를 v를 중심으로 한 v가 속한 그룹의 군집성으로 삼는다. 즉, 다음과 같이 Kv=4 일 때,



a는 b,c,d 와 directed edge로 연결되고, b는 a,c,d와 directed edge로 연결되고 … 이런 식으로 하면 Kv명의 친구들은 서로 $Kv \cdot (Kv - 1)$ 개의 directed edge로 연결되어 서로 fully connected 될 수 있다. 허나, 친구관계는 실제로 undirected edge이므로 두개의 vertices가 서로 directed edge로 연결된 것이 하나의 undirected edge로 되므로 Kv명의 친구들은 최대 $Kv \cdot (Kv - 1)/2$ 개의 직접 친구관계를 만들 수 있다.

즉, v의 Kv명의 친구들이 서로 $Kv \cdot (Kv - 1)/2$ 개의 직접 친구관계로 맺어져 있다면 이는 그 친구들이 모두 서로 다 직접 친구관계가 있다는 말이다. Cv는 실제로 Kv 명의 친구들이 서로 얼마나 친구관계를 맺고 있는가를 나타낸다.

즉, Cv = (Kv 친구들간 실제 edge 수) / ($Kv \cdot (Kv - 1)/2$) 이다. C는 모든 vertex v에 대해 구한 Cv를 평균한 값이다. 따라서, C는 어떤 네트워크가 얼마나 끼리끼리만 노는 집단성이 강한지를 나타낸다. $C \rightarrow 0$ 이면 네트워크가 집단성이 없고, $C \rightarrow 1$이면 강한 집단성이 있는 그룹들로 구성되어 있다.

- 네트워크에 총 n개의 vertices, 각 vertex당 k개의 edge가 있다고 하자.

$$n \gg k \gg \ln(n) \gg 1$$

Connected graph 에서 p →0 과 p →1의 양 끝단을 보면,

1. $L \sim n/2k \gg 1$ and $C \sim 3/4$ as $p \rightarrow 0$

   즉, p →0 되어 regular 에 가까워 질수록, 이는 C가 ¾가 되어 큰 값을 보인다는 것이고 이때 L이 n에 linear 하게 커진다. n이 vertice 개수이므로 얼마나 이 vertice를 연결하는 path의 거리가 커지겠는가?

2. $L \approx L_{\text{random}} \sim \ln(n)/\ln(k)$ and $C \approx C_{\text{random}} \sim k/n \ll 1$ as $p \rightarrow 1$

   p →1 되어 random하게 되어 네트워크가 무작위로 될수록 그 때 C 는 k/n 이 되어 매우 작은 값을 갖는다. 즉 Clustering 계수는 vertex의 평균 edge수에 비례한다.   무작위 네트워크가 되면 L 은 ln(n) 이 되어 n의 log 함수식으로 증가 (즉, 매우 느리게 증가) 한다.

✓ 그래서, regular network 같이 C (군집성)가 크면 L, 즉 노드간 평균거리가 크고

random network 같이 C가 작으면 L이 작다?   분명 양 끝단에서 이렇게 보인다.
그러나, 이것이 일반적인 것일까?   그 중간 과정에서는 어떨까?

잠깐 쉬어 가자. 수식을 음미하는 습관을 키우면 살면서 가끔씩 도움이 될 때가 있을지 모른다

앞장에 나온 수식들을 현재 우리가 알고 있는 웹과 연결해 조금 자유롭게 생각해 보면 어떨까? 많은 재미있는 생각들이 이렇게 이론을 느슨하게 적용해 보려면 시도에서 나온다.

n : 웹 있는 전체 vertices (url, 웹페이지) 수, k : vertex가 갖는 평균 edge

웹에 100억개의 url이 있다하고, 웹페이지당 평균 10개의 하이퍼링크가 있다고 가정하면 n=100억, k=10 이다.

- $n \gg k \gg \ln(n) \gg 1$

  일 경우 그래프가 보통 connected 된다고 한다. 웹에서 (n=100억) >> (k=10) 이 성립한다. 그러면 (k=10) >> ln(100억) 성립하나? ln(100억) = 23 이라 (k=10) >> ln(100억) 이 성립하지 않는다. 아… 웹이 하나의 거대한 connected component가 되려면 보다 더 random에 가까워야 하고, 페이지당 링크수가 지금보다는 더 많아야 하는 구나. 이러니, 앞서 배운 "graph structure in the web"에서 웹이 하나의 거대한 connected component가 아닌 것이 이해가 된다.
  이처럼 모든 것이 통한다.

$L \sim n/2k \gg 1$ and $C \sim 3/4$ as $p \rightarrow 0$

- 일단, C가 ¾에 가깝다는 것은 C 값이 매우 크다는 말이다. 이론적으로 C가 최대 1이므로, 이 값은 큰 값이다. 즉, 웹이 이렇다면 웹은 많은 그룹들로 구성되어있고, 이 그룹들의 구성원 url 들은 자기들끼리 매우 강하게 결속, 즉 내부 링크들로 매우 밀도있게 연결되어 있다는 말이다. 개별 url 관점에서 보면, 내가 링크한 url들은, 자기들끼리도 서로 링크들을 많이 한다는 얘기다.

  이럴 경우 L이 n/2k 에 접근하고 이 값은 1 보다는 매우 크다고 한다. 잠깐 여기서 L을 생각해 보자. 웹에서 L은 임의의 두 url간 최단거리의 평균이다. 이 수식은 웹에서 C가 매우 커지면, 두 웹페이지들을 링크로 연결하는 거리가 n에 비례하고, k에 반비례한다는 말이다. 직관적으로 말이 된다. 웹이 가까이 있는 URL끼리만 노는 끼리끼리 문화가 커지면, 웹의 웹페이지수가 늘수록 그 크기에 비례해 노드간 거리가 커진다는 얘기다. 각 노드당 링크수가 커지면 거리가 작아지지만 실제로 웹에서 개별 페이지당 담고있는 링크 개수가 웹이 커질 수록 많아진다고 생각하기 어려우니 실제로 k 값은 거의 상수라고 생각하면 좋겠고 따라서 L은 그냥 n에 비례한다고 생각하는 것이 편하다.

Granovetter의 “strength of weak ties” 를 연결해 보자. C가 크다는 말은 cliqueness가 크다고 할 수 있다. 또, Clique간을 연결하는 weak tie가 존재할 수 있지만, 그 tie도 지역간 가까운 clique를 연결하는 것이라 생각할 수 있다. (Granovetter의 논문은 그래프의 topology측면은 별로 고려하지 않았다). 웹에서 링크로 연결되어 가깝다는 것은 것은 지역/지리적인 특성을 타는가? 아니면, 웹페이지 내용의 밀접성을 더 타는가? 한글로 된 url은 아무래도 다른 한글 url을 링크하니 지역성을 탄다고 해야 하나?

$$L \approx L_{\text{random}} \sim \ln(n)/\ln(k) \text{ and } C \approx C_{\text{random}} \sim k/n \ll 1 \text{ as } p \rightarrow 1$$

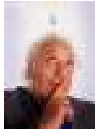
- 그래프의 edge들이 제멋대로, 생각나는 대로, 무작위로 연결될 때의 L과 C를 나타낸 것이다.

우선, 웹은 무작위 그래프가 아니다. 그 비슷한 것이, 웹의 한 구석에서, 우리가 모르게 스멀스멀 성장할 가능성도 별로 없다. Why? 어떤 웹페이지 저자가 아무 관련도 없는 웹페이지를 링크로 연결할까? 있다면, 극히 소수이고 아마 또라이일것이다.

웹이 무작위 비슷하다고 그냥 가정하자. 이 경우 C 는 k/n 에 접근한다 한다. 앞서, k는 웹의 크기와 관계없이 그냥 고정값일 가능성이 많다고 했다. 이러니, 웹이 무작위하게 될 수록 n에 반비례하여 C값이 작아진다는 것이다. K=10, n=100억 하면 C 값 매우 작다. 이미 작은 값이므로 웹이 성장해 n이 더욱 커져도 C 값의 크기는 별로 영향 받을 것 같지도 않다.

핵심은 L이 ln(n)/ln(k) 라는 것이다. 100 > k >1 이어서 k가 고정값에 가까우니 ln(k)도 고정값이고, 그러니 L이 ln(n)에 접근하는 것이 식의 핵심이다. Log를 적용하면 어떤 값이 얼마나 크게 변하건, 그 증가속도가 완전히 밋밋해진다. 10^{100}이면 아마 우주의 원자수 보다 큰 수일지도 모르는데, 여기에 log를 적용하여 $\log_{10}(10^{100})$ 이면 100 이 되어 그저 그런 값이 되어버리고 만다. 따라서 이 얘기는 웹이 아무리 커져도, 웹의 구조가 무작위 네트워크 같다면, 임의의 두 url간의 거리가 기껏해야 수십개의 link를 따라가면 닿을 수 있다는 말이다. n=100억이면 L ~ ln(100억)/ln(10)=23/2.3 = 10 이다. 웹이 무작위라면 url간 평균거리가 10이다.

웹이 무작위그래프가 아니라고 했다. 그렇다고 규칙적이지도 않다. **Broder** 등이 쓴 **"graph structure …"** 에서 우리는 웹의 실제 구조를 보았다. 무슨 얘기인가 하면, 웹에서 **2**억개의 **vertices** 취해 모델을 만들어보니 임의의 두 **url**간에 **path**가 있는 확률이 **0.24**이고 **path**가 있다면 그 때 **L**이 약 **16**이라 했다. 참고로 웹이 무작위그래프라 가정해 **L**에 **ln(n)/ln(k)**를 적용하면 **19.11/2.3**이 되어 **8.3**이다. 즉, **Broder**가 본 **16**보다는 작지만 그렇다고 웹이 **Regular** 한 것보다는 훨씬 더 무작위에 가까운 값이다. 어떻게 보아야하나? 여기에 이 논문의 핵심내용이 빛을 보인다.

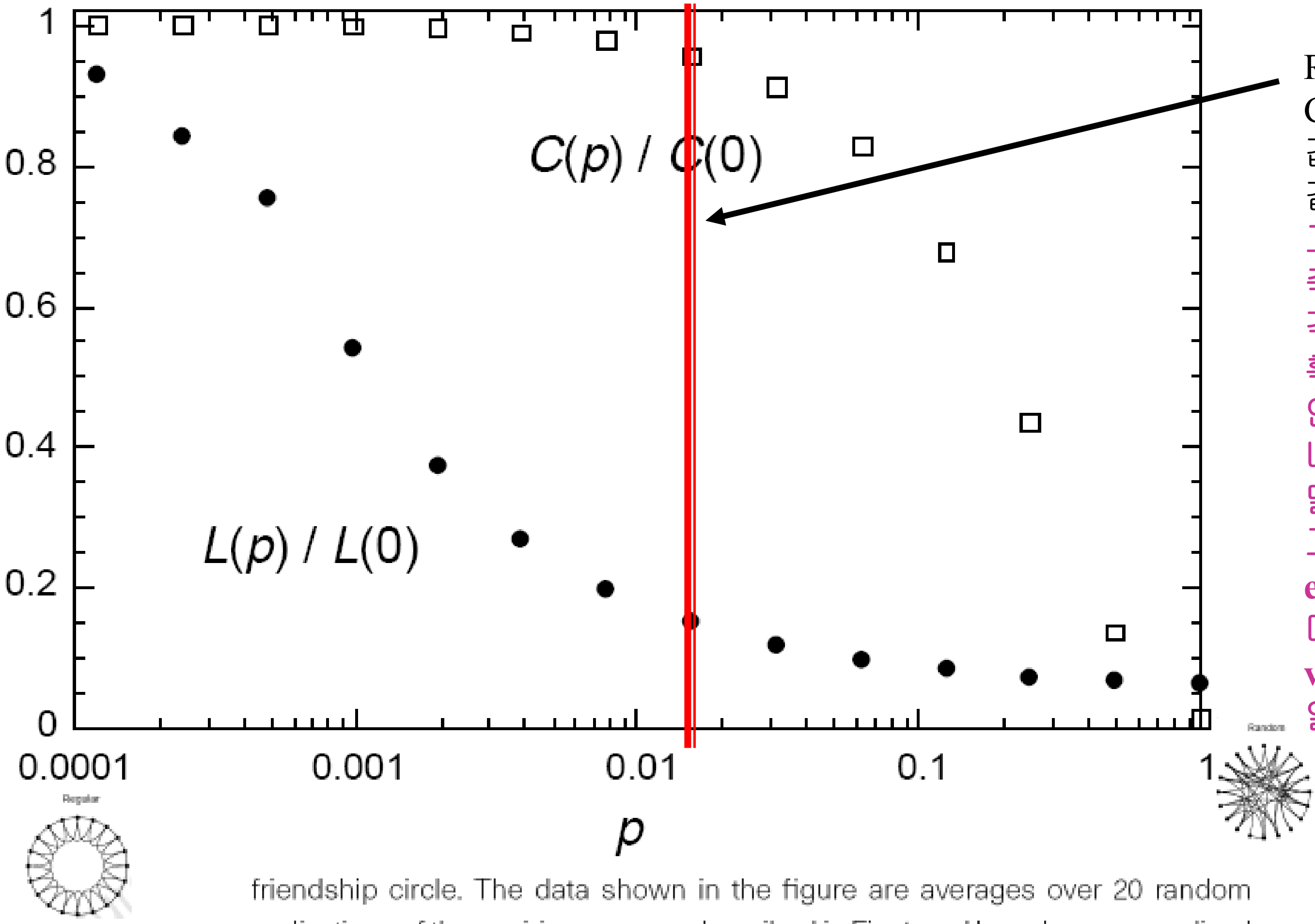


Cream of Watts & Strogatz 의 얘기

- **Watts와 Strogatz는 clustering 정도가 매우 높은 regular 네트워크에서 노드간 평균거리가 매우 클 때, 작은 수의 edge를 무작위로 먼거리에 있는 다른 cluster에 있는 vertices 들에 연결만 해도 노드간 평균거리가 급격히 작아진다는 것을 보였다. 그래도, 그 네트워크는 여전히 전체적인 clustering 정도가 크다. 대부분의 개별 클러스터 형상은 변하지 않았고, 대부분 node들의 edge 관계도 변하지 않았다. 즉, 개별 노드의 관점에서 자기가 속한 클러스터의 변화가 느껴지지 않으며, 전체적인 네트워크의 변화는 전혀 감도 잡지 못한다. 개별 노드가 보는 미시적인 관점에서 clustering 변화를 감지하지 못하는 가운데 몇 개의 무작위 링크의 도입으로 순식간에 거시적인 지표, 즉 노드간 거리가 급격히 변한다. 대부분의 개별 노드는 모르지만 (물론 당사자는 알지만), 전체 네트워크 특성은 변했다.**

"On the contrary, Fig. 2 reveals that there is a broad interval of *p* over which *L*(*p*) is almost as small as *L*random yet $C(p) \gg C_{\text{random}}$. These small-world networks result from the immediate drop in *L*(*p*) caused by the introduction of a few long-range edges. Such 'short cuts' connect vertices that would otherwise be much farther apart than *L*random. For small *p*, each short cut has a highly nonlinear effect on *L*, contracting the distance not just between the pair of vertices that it connects, but between their immediate neighbourhoods, neighbourhoods of neighbourhoods and so on. By contrast, an edge

removed from a clustered neighbourhood to make a short cut has, at most, a linear effect on *C*; hence *C*(*p*) remains practically unchanged for small *p* even though *L*(*p*) drops rapidly. The important implication here is that at the local level (as reflected by *C*(*p*)), the transition to a small world is almost undetectable. To check the robustness of these results, we have tested many different types of initial regular graphs, as well as different algorithms for random rewiring, and all give qualitatively similar results. The only requirement is that the rewired edges must typically connect vertices that would otherwise be much farther apart than *L*random."



Regular 그래프에서와 같이 C 값이 여전히 큰데, 벌써 L 값은 random 그래프에서와 같이 L 값이 많이 작아졌다. **그런데 이것이 바로 p=0.02, 즉 regular 그래프에서의 각 edge를 0.02의 낮은 확률로 다른 vertex로 연결했을 경우에도 나타났다는 것이다. 다른 말로, regular 그래프에서의 약 2% edge들만 이런 파격을 부려 다른 클러스터에 있는 vertices로 연결될 때 이런 일이 생긴다는 말이다.**

friendship circle. The data shown in the figure are averages over 20 random realizations of the rewiring process described in Fig. 1, and have been normalized by the values $L(0)$, $C(0)$ for a regular lattice. All the graphs have $n = 1,000$ vertices and an average degree of $k = 10$ edges per vertex. We note that a logarithmic horizontal scale has been used to resolve the rapid drop in $L(p)$, corresponding to the onset of the small-world phenomenon. During this drop, $C(p)$ remains almost constant at its value for the regular lattice, indicating that the transition to a small world is almost undetectable at the local level.

Small world 특성이 social network, 송전망, 생물의 신경망과 같은 보편적인 네트워크에서 발견된다.

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Characteristic path length L and clustering coefficient C for three real networks, compared to random graphs with the same number of vertices (n) and average number of edges per vertex (k). (Actors: $n = 225{,}226$, $k = 61$. Power grid: $n = 4{,}941$, $k = 2.67$. *C. elegans*: $n = 282$, $k = 14$.) The graphs are defined as follows. Two actors are joined by an edge if they have acted in a film together. We restrict attention to the giant connected component[16] of this graph, which includes ~90% of all actors listed in the Internet Movie Database (available at http://us.imdb.com), as of April 1997. For the power grid, vertices represent generators, transformers and substations, and edges represent high-voltage transmission lines between them. For *C. elegans*, an edge joins two neurons if they are connected by either a synapse or a gap junction. We treat all edges as undirected and unweighted, and all vertices as identical, recognizing that these are crude approximations. All three networks show the small-world phenomenon: $L \gtrsim L_{\mathrm{random}}$ but $C \gg C_{\mathrm{random}}$.

Small world 네트워크 특성이 dynamical 시스템에서도 나타난다.

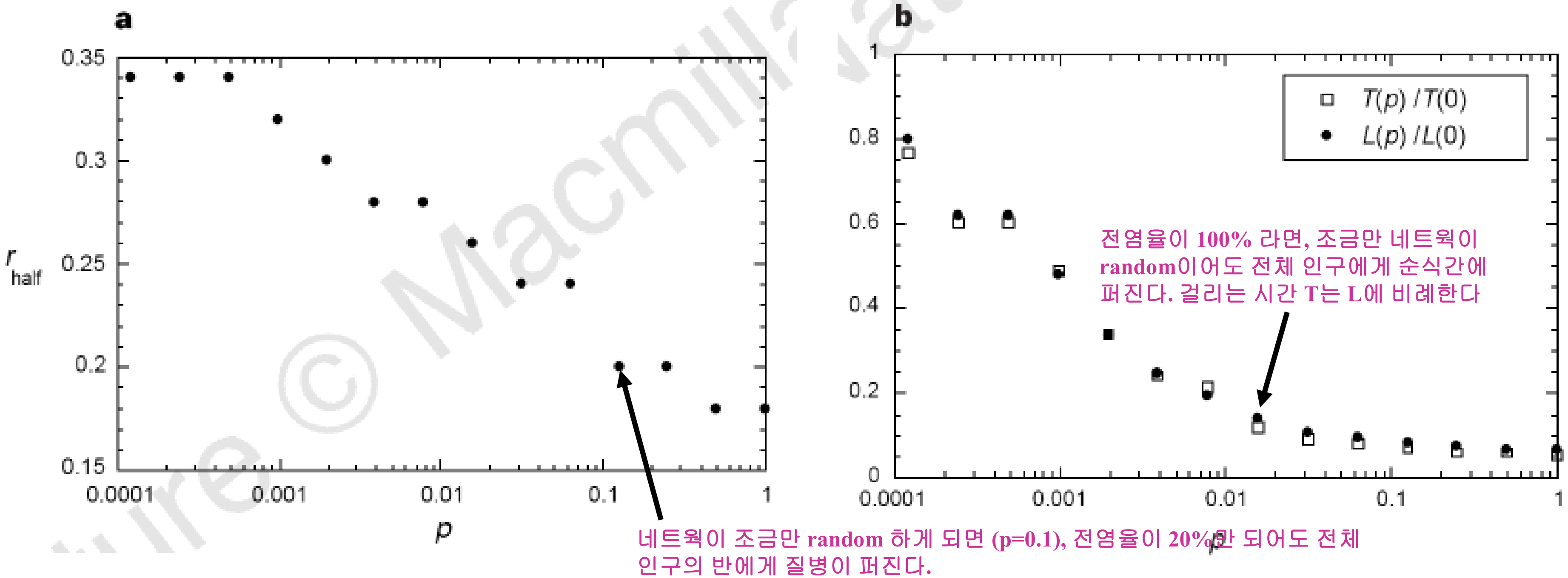
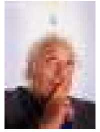


Figure 3 Simulation results for a simple model of disease spreading. The community structure is given by one realization of the family of randomly rewired graphs used in Fig. 1. **a**, Critical infectiousness r_{half}, at which the disease infects half the population, decreases with p. **b**, The time $T(p)$ required for a maximally infectious disease ($r = 1$) to spread throughout the entire population has essentially the same functional form as the characteristic path length $L(p)$. Even if only a few per cent of the edges in the original lattice are randomly rewired, the time to global infection is nearly as short as for a random graph.



음미하기

음.

우리 인간관계에서, 사회에서, 웹에서 관계들이 regular 하다는 말은 어떤 의미로 해석할 수 있을까? 관계가 random 하다는 얘기는? 아니, random과 관계라는 말이 서로 모순되는 말이 아닌가? 관계라는 것은 그런 관계 패턴이 나타날 만큼의 기간, 성장이 필요한 것이 아닌가?

Regular와 Random 두 개 중에서 앞서의 관계 중 굳지 한 개를 고르라면 어떤 것을 고를까? Random, 그저 개념없이, 부딪히는대로 관계가 생긴다는 말이지? 특정한 경우에, 특정한 사람중에 이런 일이 생길 수도 있겠지만 별로 보편적이라 생각되지 않는다. 우리들은 우리 주변에, 비슷한 사람끼리 관계를 맺을 가능성이 많기에, 실제 관계들은 regular에 더 가깝다.

L을 정보의 전파, 습득, 공유와 연결해 생각해 보자. L이 클수록, 정보의 전달이 시간걸리고, 왜곡되고, 끊어지기 쉬울 것이다. L이 작다는 것은 전체 네트웍이 서로 매우 긴밀히 연결되었다는 것인데, 실제로 우리 세상은 그렇지 않다. 섬들이 다리로 느슨히 연결되어 있다. 대부분의 사람들은 섬에 있고, 모든 사람들이 다른 모든 사람들과 직접 연결되기는 가능하지 않고 소수의 다리를 거쳐 연결된다.

Stability와 flexibility를 도입해 보자. C가 클수록, 개별 클러스터는 안정성이 좋지만 외부로의, 전체적인 변화에 따른 flexibility는 약해진다. Small world network가 보인대로 소수의 파격 링크가 전체 정보전달, flexibility 향상에 큰 역할을 한다. 그런데, 이런 edge들은 Granovetter에 의하면 weak edge이다. 어떻게 생성되고, 유지되고, 사라질까? 이들을 통합적으로 볼 수 있을까?