SOSC 4300/5500: Small-world Networks, Weak Ties and Diffusion

Han Zhang

Nov 17, 2020

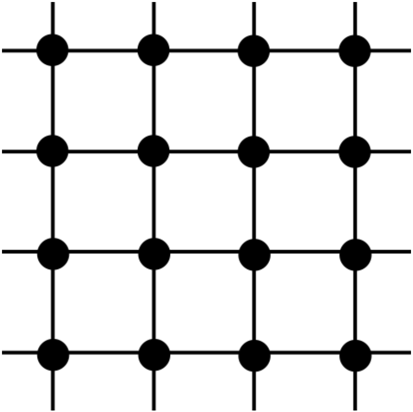
Outline

Theoretical modeling of small-world networks

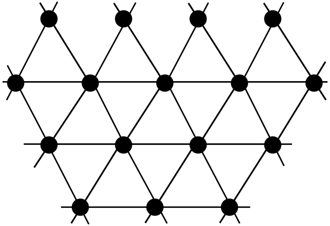
- So far, we have taken an empirical approach toward understanding small-world networks
  - e.g., measuring/describing the diameter
- A theoretical approach, however, asks what's the underlying conditions that produce small-world networks
- Before answering these questions, we first look at two simple network examples that are not small-world networks

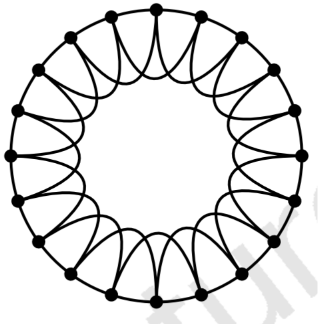
Simplest network: regular network

- Every node has the exact same number of edges



Simplest network: regular network



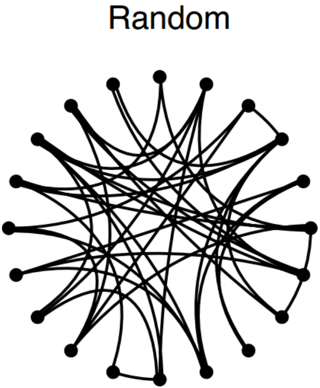


Regular networks are not small-world networks

- Why?
- Diameter is large

Simplest network model: Erdos-Renyi network

- Erdos-Renyi network: the simplest random network
- N nodes
- Each node has a probability to make an edge with any other with a probability p
- $N \cdot p$ edges in expectation



Erdos-Renyi network is not small-world network

- `http://www.networkpages.nl/CustomMedia/Animations/RandomGraph/ERRG/AddoneEdgepATime.html`
- Is this network small world? No
- The diameter is small enough
- But the clustering coefficient $\to 0$ when N increases

Comparisons

- The two ideal types, regular networks and random networks, looks very different
- And they are all different from small-world networks

	Diameter L	Clustering Coefficient C
Random	small	small
Regular	large	large
Small-world	small, around $log(N)$	large

Small-world Phenomena beyond social networks

- “the small-world phenomenon is not merely a curiosity of social networks, nor an artefact of an idealized modelit is probably generic for many large, sparse networks found in nature”

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

Diameter vs. Clustering Coefficients

- More on diameter vs. clustering coefficients
- Diameter is a global measure
  - It's the average shortest distance between each pairs of nodes
- Local clustering coefficient is a local measure
  - You can collect more complete information about an indidividual, by asking whether two of his friends know each other
- Local clustering coefficient is easier to measure than diameter

Watts-Strogatz Model

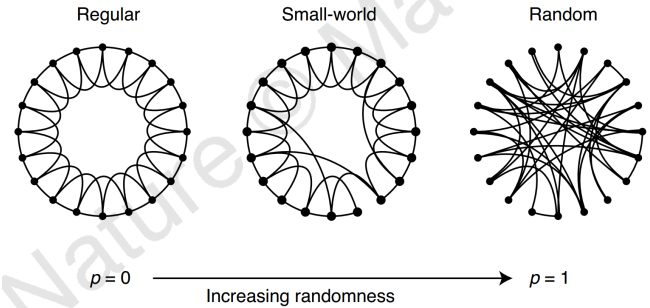
- Duncan J. Watts and Steven H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature **393** (1998), no. 6684, 440–442
- Perhaps the most influential work of modern network analyis
- Key intuition: only adding several long-range edges can turn regular network into a small-world network
- Why? These long-range edges connect otherwise distance nodes

Watts-Strogatz Model

- Theory building from simulation, or agent-based modeling
  - Start from a regular network
  - "We choose a vertex and the edge that connects it to its nearest neighbour in a clockwise sense. With probability p, we reconnect this edge to a vertex chosen uniformly at random over the entire ring, with duplicate edges forbidden; otherwise we leave the edge in place."
  - Increasing p makes the graph more random
  - $p = 1$ makes the network completely random
- Demo: `http://www.netlogoweb.org/launch`

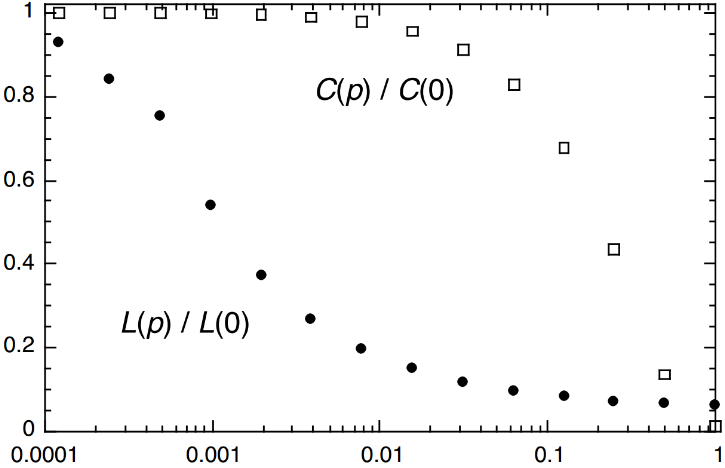
Watts-Strogatz Model

- Key finding from Watts and Strogatz:
  - a very small number of p would suffice to turn a regular network into small-world network



Watts-Strogatz Model

- A small p leads to a small-world network
  - Large clustering coefficient $C(p)$
  - Small diameter $L(p)$



Strength of Weak Ties

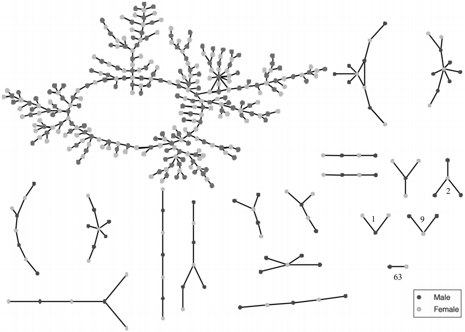
- Watts-Strogatz model nicely connects with the study of weak ties
- Some intellectual history
- 1960s: Milgram's small world experiments
    - Hinted that long-range ties are important for bridging otherwise distant nodes
- Mark S. Granovetter, *The Strength of Weak Ties*, American Journal of Sociology **78** (1973), no. 6, 1360–1380
    - Substantive question: *where do people get useful information during job searches*
    - Inspired by Milgram: from weak ties
    - Granovetter did not touch the idea of small-worlds
- Watts-Strogatz model links the idea of weak ties with the idea of small worlds

Strength of Weak Ties

- Granovetter's study has two parts
- The first part is a mathematical theory of why weak ties are important in spreading information
  - This part is very general
- The second part is an empirical application in job search settings
- Funny story: this article was also rejected early on
- `https://scatter.files.wordpress.com/2014/10/granovetter-rejection.pdf`

Connectivity of networks

- Connectivity of networks
- If you can go from one node to any other node in a network, the network is connected
- Otherwise, each subgraph that is connected inside is called connected component
- And the largest connected subgraph is called largest connected component

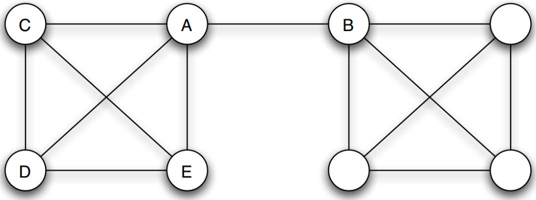


Connectivity

- Theorem (Erdos and Renyi, 1960): in a Erdos-Renyi random network
  - if $p > \frac{1}{n}$, then almost surely the largest connected component will contain over $n^{\frac{2}{3}}$ nodes
    - in human language, at least $n^{\frac{2}{3}}$ nodes will be connected
  - if $p > \frac{ln(n)}{n}$, then almost surely the entire network is connected
- Implication: it's really easy for a social network to be connected
- And the larger the size n, the easier for a network to be connected

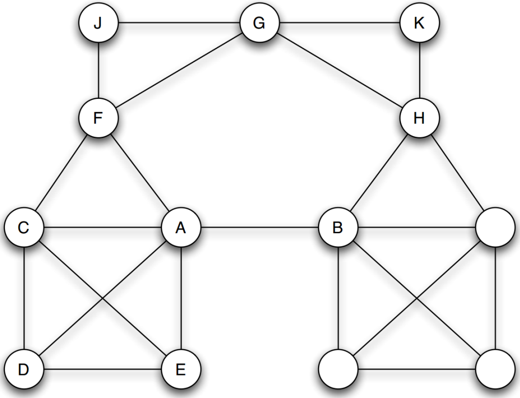
Bridge

- Bridge characterizes key edges in reducing the diameter of networks
- An edge $< A, B >$ is an bridge, if its removal will make A and B into two separate connected components
- Another way to think: the removal of a bridge increases the distance to ∞



Local Bridge

- In large networks it is rare to see an bridge
- A weaker version is local bridge: $< A, B >$ is a local bridge if they do not have common friends
    - In other words, the removal of a local bridge $< A, B >$ will increase the distance between A and B to at least 2



Math model of weak ties

- Now let us link local bridge with weak ties
- Simplest network model: 0/1
- Adding a little bit complexity: 0 (no tie)/strong tie / weak tie
- And adding a critical assumption: strong triadic closure property

if A have strong ties to both B and C, then B and C must have at least a weak tie

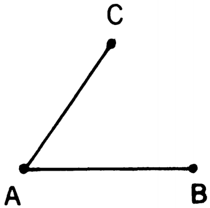


FIG. 1.—Forbidden triad

Weak ties and Local Bridge

- Theorem: (Granovetter, 1973)

  For any node A in a network, assuming:

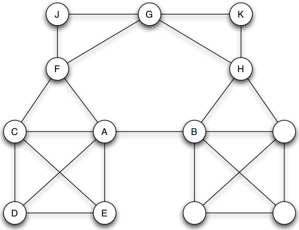
  1. *A follows strong triadic closure property*
  2. *A have at least two strong ties*

  then local bridges that include A must be weak ties

- Note: weak ties are not necessarily local bridges

Proof

- *Proof by contradiction*
  - Assume $< A, B >$ is a local bridge and a strong tie (contradiction here)
  - Since A is involved in at least two strong ties by Assumption 2, and the edge to B is only one of them, it must have a strong tie to some other node
  - Say that other node is F
  - By Assumption 1 (strong triadic closure property), B and F must be friends
  - This violates the definition of local bridge: A and B should not have common friends



Contributions of Granovetter's model

- Before Granovetter, weak tie were rarely studied
  - Think about survey of ego-networks, which almost always study *strong ties*, not weak ties
    - E.g., GSS: "Whom do you discuss important matters with"
- Granovetter's contribution:
  - One extra piece of complexity (from 0/1 to 0/weak tie/strong tie)
  - And one plausible assumption (strong triadic closure)
  - Lead to an insightful observation: weak ties are very important because they must be local bridges, which reduces distances between nodes and thus makes information diffusion faster

Comparisons

	Focus	Empirical/Theoretical
Milgram	Small World	Empirical; measurement
Granovetter	Weak Ties	Both
Watts-Strogatz	Small world and weak ties	Theoretical

Empirical part of Granovetter 1973

- Granovetter then tested his argument in real-world empirically
- Interviewed professionals in Boston who found a job through a contact, asked these people:“ how often did they see the contact?”
  - often (at least twice a week): 17%
  - occasionally (more than once a year but less than twice a week): 56%
  - rarely (once a year or less): 28%
- What are potential problems of this approach?

Shortcomings

- Selecting on dependent variables. Maybe people find jobs through weak ties because they have more weak ties?
  - Valery Yakubovich, *Weak Ties, Information, and Influence: How Workers Find Jobs in a Local Russian Labor Market*, American Sociological Review **70** (2005), no. 3, 408–421
- Context: ties may be more important in finding jobs in other countries strong
  - Yanjie Bian, *Bringing Strong Ties Back in: Indirect Ties, Network Bridges, and Job Searches in China*, American Sociological Review **62** (1997), no. 3, 366–385

Modern advancements

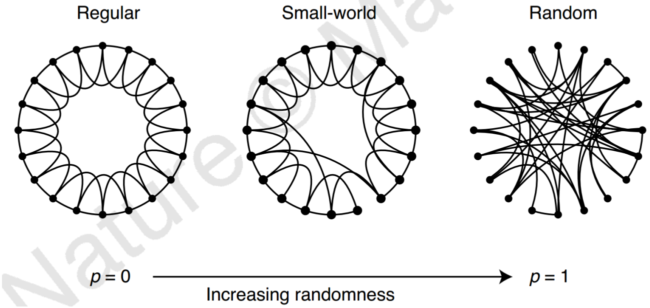
- Eric Gilbert and Karrie Karahalios, *Predicting tie strength with social media*, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA), CHI '09, Association for Computing Machinery, 2009, pp. 211–220
- Using Facebook's data (objective, behavioral) to predict tie strength (subjective, answered by respondents)
- Predictors: 74 different measures constructed by the authors
  - [in class activities]: write down a measure of tie strength based on Facebook data
- Outcomes: "how strong is your relationship with this person?"
  - There are other subsequent outcomes, such as "how helpful would this person be if you were looking for a job"

Results

Top 15 Predictive Variables	β	F	p-value
Days since last communication	-0.76	453	< **0.001**
Days since first communication	0.755	7.55	< **0.001**
Intimacy × Structural	0.4	12.37	< **0.001**
Wall words exchanged	0.299	11.51	< **0.001**
Mean strength of mutual friends	0.257	188.2	< **0.001**
Educational difference	-0.22	29.72	< **0.001**
Structural × Structural	0.195	12.41	< **0.001**
Reciprocal Serv. × Reciprocal Serv.	-0.19	14.4	< **0.001**
Participant-initiated wall posts	0.146	119.7	< **0.001**
Inbox thread depth	-0.14	1.09	0.29
Participant's number of friends	-0.14	30.34	< **0.001**

Nonlinearity in networks

- Watts-Strogatz model also reveals non-linearity in networks
- Increasing p from 0 (more re-wired edges)
  - has a highly nonlinear negative effect on diameter L
  - has a linear negative effect on clustering coefficient C
- Implications: "equally significant changes in global structure can result from changes in local structure that are so minute as to be effectively undetectable at the local level."



Nonlinearity in networks

- This is an example of the "butterfly effect"
  - Also called phase transition in academic jargon
- A small change around a tipping points (here, $p = 0$ or $p = 0.001$) for some individuals
  - Lead to a drastically different world for the entire network
- These kinds of phase transitions cannot be captured by linear regressions and its extensions
- Observing many other properties in networks exhibit phase transitions, network scientists generally doubts the linear way of thinking about the world
  - Instead, they want to study connections as a whole, or as a "complex system"

Social diffusion/contagion

- Let us look at another example that exhibits phase transition
- Social diffusion/contagion: individual behaviors can spread along social networks
- Early adopters will make some individual decisions
- Examples
  - Participating a protest
  - Start using an innovative product
  - Know some new information
  - Infected diseases
  - Leaving parties
- Other people's behaviors depend on how they are connected with early adopters

Consequence of social diffusion

- Assume some early adopters have adopted certain behaviors, such as spreading some rumors or starting using a new iPhone app.
  - In academic jargon, these early adopters are activated
- Whether the entire network will adopt that behavior?
- Or ask it differently, how quick the entire network will adopt that behavior?

Watts-Strogatz model's implication

- As long as there is a few weak ties, network becomes a small-world
- And small-world network means that information and disease can spread very rapidly
- Bad side: it takes only a few contagious people to travel between remote regions to make the entire population highly vulnerable to epidemics
- Good side: only a few weak ties can help you obtain the critical information about job searches (Granovetter)
- But overall, as long as $p > 0$, it's very quick for the entire network to adopt a new behavior

Diffusion on Watts-Strogatz's small-world networks

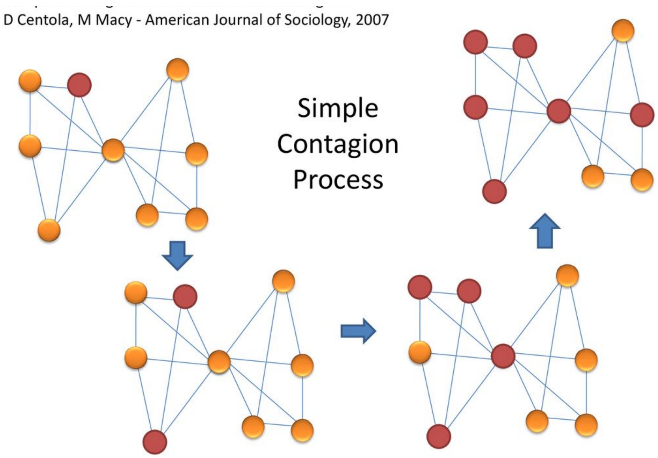
- demo:

http://modelingcommons.org/browse/one_model/5216#model_tabs_browse_nlw

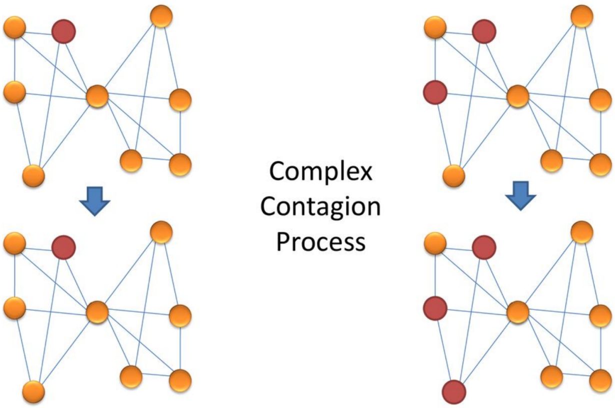
Simple vs. Complex Diffusion

- However, there are two types of diffusion
- Simple contagion: e.g., the spread of COVID or spread of information
  - As long as you have 1 close contact infected, you have a chance to be infected
  - Of course, if you have more than 1 close contacts infected, your chance of infection is a lot more higher
- Complex contagion: e.g., joining a risky protest
  - One friend participating may not be enough
  - Perhaps need multiple friends' confirmation
- [In-class discussions]: can you think of more examples?

Simple contagion



Complex contagion



Complex contagion

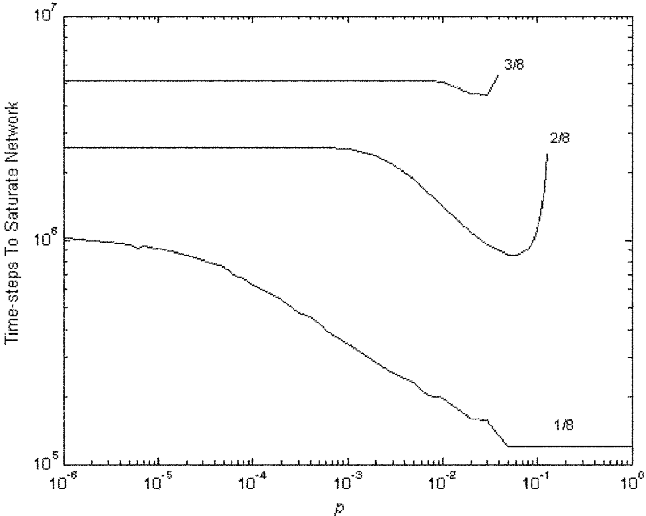
- Damon Centola and Michael Macy, *Complex contagions and the weakness of long ties*, American Journal of Sociology **113** (2007), no. 3, 702–734
- The exact setup as Duncan J. Watts and Steven H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature **393** (1998), no. 6684, 440–442
- But add another parameter: threshold
  - Threshold $= 1$: simple contagion; if one of your friends adopt something, you have a non-zero chance to adopt that
  - Threshold $= k > 1$: complex contagion; you will adopt something only if no less than k of your friends adopted that
- Overall, the model has two critical parameters now
  - p: the rewiring probability; 0 vs > 0 (determining the number of weak ties)
  - threshold: 1 vs. more than 1

Complex contagion

- If threshold > 1, then there are two competing forces as p increases (adding weak ties)
- Diameter quickly decreases; good for diffusion
- But a weak tie will certainly make the remote node not satisfying the threshold requirement; bad for diffusion
- Macy and Centola found that increasing p still helps a little bit for quick diffusion to the entire network, but not that much as compared with simple contagions

Complex contagion

- Look at the NetLogo demo again
- Also Macy and Centola's figure



Complex contagion

- Conclusions:
- Still, there is a critical threshold for p, but it's no longer 0
- And the range of p that allows for speedy diffusion is very narrow, compared with the case of simple contagion
- Finally, the effect of p (adding weak ties) on diffusion outcomes for the entire network is also non-linear

Summary

- For simple contagion, as long as there is a few ($p > 0$) weak ties, the diffusion speed to the entire network is drastically reduced; the effect of weak ties on diffusion speed is non-linear
- For complex contagion; the effect of weak ties on diffusion speed is also non-linear, but the tipping point for p is no long 0.
- Either way, if you measure individual's tie strength and want to use linear regressions to study global diffusion patterns, it's likely to fail

Two types of computational social sciences

- Two parallel developments of computational social sciences
- For studying complex networks
  - Social phenomena are non-linear; we need to study it as a complex network
  - A natural hybrid of theory-driven mathematical simulations and empirical analysis using big data
  - A new paradigm; from studying attributes to studying connections; big mind shift.
- For measurement
  - E.g., applying machine learning techniques on text data to generate some variables, and then put these variables into a linear regressions to test some theories
  - Mostly an empirical approach: old theory + new data