Lecture 11: Strategic Aspects of Diffusion and Contagion

Alexander Wolitzky

MIT

6.207/14.15: Networks, Spring 2022

Strategic Aspects of Diffusion and Contagion

Diffusion processes covered so far have been "mechanical."

- The process runs on its own without any interventions by outsiders or strategic decisions by the individuals/nodes.

In reality, interventions and strategic decisions play key roles in many diffusion processes.

- Social distancing, lockdowns, or vaccinations to **slow or stop** a pandemic.
- Advertising or "seeding" to **speed up** adoption of a new product or technology.
- **Individual decisions** about whether to adopt a product or technology based on what one's neighbors are doing.

Plan

1. Interventions to hinder diffusion: targeted removal, endogenous activity ("social distancing").
2. Interventions to help diffusion: optimal seeding.
3. Strategic contagion: spread of coordination behaviors.

Hindering Diffusion: Random Node Removal

As we've seen, in the ER or SIR model, randomly vaccinating fraction π of the population reduces the expected number of meetings with unvaccinated individuals to $R_0 (1 - \pi)$.

This reduces the share of unvaccinated people who ever get sick (i.e., the size of the giant component) to the solution to $R(\infty) = 1 - e^{-R_0(1-\pi)R(\infty)}$.

- The total share of people who ever get sick is $(1 - \pi) R(\infty)$.

Hindering Diffusion: Targeted Node Removal

Targeted removal can do even better.

A simple model: removing fraction π of the population with the highest degrees.

This reduces the share of people who ever get sick by more than randomly removing the same fraction.

How much greater the reduction is depends on the degree distribution.

We'll skip the calculation, but in general the more heterogeneous is the degree distribution, the greater the gain from removing the highest-degree nodes rather than random nodes.

This difference is especially important for scale-free graphs ($P(d) = cd^{-\gamma}$), which have very skewed degree distributions.

Application: Is the Internet Robust?

Recall that web links are distributed approximately power-law with $\gamma < 3$ (most estimates say $\gamma \in (2.1, 2.7)$).

We have seen that scale-free networks are robust to random node removal: randomly remove 99% of the links and there will still be a giant component (i.e., the contagion threshold is $\pi^* = 1$).

- This is a good model for individuals who happen to be immune to a virus, or for webpages that randomly go down.

However, the presence of the giant component comes from the presence of a few very high-degree nodes serve as "hubs."

- Removing the 3% highest-degree nodes destroys the giant component.
- This could be a good model for a cyberattack.

A catchphrase for this phenomenon: "The Internet is robust, yet fragile."

Hindering Diffusion: Endogenous Activity

In the basic SIR model, people can't do anything to reduce their risk of infection.

- $\dot{I}(t) = \beta S(t) I(t)$, where β is an exogenous parameter.

In reality, people can take actions to reduce infection risk, although these actions typically involve some costs.

- Masking, social distancing, etc..
- Especially important for very dangerous diseases like Covid-19.

Paying a cost to reduce your infection risk ("your β") in period t is worth it if $I(t)$ is high enough.

At the population level, this causes in a reduction in β when $I(t)$ is high.

- Individual behavioral responses endogenously flatten the curve to some extent.

Hindering Diffusion: Endogenous Activity (cntd.)

This type of **behavioral SIR** model fits the data from the Covid-19 pandemic better than the standard SIR model.

- Cases typically rose fast but slower than exponentially, reflecting people and governments taking precautions.

Several variant of behavioral SIR models have been developed (especially since Covid).

We'll present a simple one that gives the main idea.

A Behavioral SIR Model

Suppose that at each point in time every individual can pay a **cost** c to completely eliminate her risk of getting infected at that time.

- Let's call this action **"vigilance."**

Suppose individuals perceive the **harm** they suffer if they get infected as h.

Since susceptible individuals get sick at rate $\beta I(t)$ if they are not vigilant:

1. If $I(t) < I^* = \frac{c}{h\beta}$, no one if vigilant.
2. If $I(t) > I^*$, everyone is vigilant.
3. If $I(t) = I^*$, each individual is indifferent between being vigilant or not being vigilant, so it may be that some people are vigilant while others are not.
    - When we introduce game theory language in the 2nd half of the class, we will call such a situation a **mixed-strategy equilibrium**.

Behavioral SIR Model (cntd.)

Under our simplifying assumption that being vigilant completely eliminates risk, $I(t)$ can never actually rise above I^*.

- If it did, everyone would be vigilant, so $\dot{I}(t)$ would immediately fall to $-\gamma I(t)$, and $I(t)$ would drift back down to I^*.

Behavioral SIR Model (cntd.)

So, the epidemic will proceed in 3 phases:

1. **Rising phase:** Initially, $I(t)$ is close to 0, so no one is vigilant and $I(t)$ starts rising. In this phase, the epidemic proceeds exactly as in the standard SIR model.
2. **Plateau phase:** Once $I(t)$ hits I^*, enough people start being vigilant so that $I(t)$ remains **exactly** at I^* for an extended period of time. This is the only possibility because
   - $I(t)$ can't rise above I^* (everyone would start being vigilant)
   - $I(t)$ can't fall below I^* before herd immunity is reached (everyone would stop being vigilant, and $I(t)$ would rise again)
   - (Aside: This is an example of a game where the *only* equilibrium involves mixed strategies. This is typical of situations where people want to do the opposite of others, like a pandemic where I can relax if others are vigilant but I should be vigilant if others relax.)
3. **Declining phase:** Once herd immunity is reached ($S(t) = 1/R_0 = \gamma/\beta$), everyone stops being vigilant, and $I(t)$ gradually falls to 0 as in the standard SIR model.

Behavioral SIR Model: Comments

This simple model has several features that reflect the Covid-19 pandemic pretty well.

1. There is no single huge spike in cases (contrary to SIR-based predictions like the Imperial College model).
2. But the pandemic lasts at a relatively high level for a long time, and it takes a long time to reach herd immunity (again contrary to SIR-based predictions, which suggested a more intense but shorter epidemic).
3. The long-run share who get sick, $R(\infty)$, is slightly lower in the behavioral SIR model than the standard SIR model. In both models the same herd immunity threshold is ultimately reached and there is some overshooting, but there is less overshooting in the behavioral SIR model because $I(t)$ is lower at the point where herd immunity is reached.
    - Depending on the parameters, this effect can be small, and the difference between behavioral SIR and standard SIR can be more about spreading infections over time than reducing the total number of infections.

Helping Diffusion: Optimal Seeding

We want to stop bad diffusions like epidemics, but we want to help spread good diffusions like beneficial new products or technologies.

The flip-side of optimal removal/vaccination is **optimal seeding**: which nodes should we "seed" to try to maximize the spread of the diffusion?

- "Viral marketers" decide whom to initially inform about their products.
- Development economists decide which villagers to inform about a new opportunity or technology.
- Hackers decide where to insert viral malware. (The diffusion is bad for society, but good from the hackers' perspective.)

The Diffusion of Microfinance

An influential paper by economists Banerjee, Chandrasekhar, Duflo, and Jackson (2013) takes this perspective to study the diffusion of information about the expansion of a microfinance program in a group of Indian villages.

- The microfinance institution entered each village by inviting a group of individuals to a meeting, where they were told about the expansion.
- The economists collected survey data about the social network in the village and about who ultimately took up the microfinance program.
- They asked how the network position of the "seeds" affected ultimate take-up in the village.
- They show that a measure of the seeds' centrality (similar to Katz-Bonacich centrality) is a strong predictor of village take-up (much stronger than simpler measures like degree).
- They conclude that who gets seeded matters for diffusion in this context, and that centrality is a key consideration.

Just a Few Seeds More?

A new paper by Akbarpour, Malladi, and Saberi (2021) questions the importance of seeding, through an interesting argument: do we really need to seed central nodes, given that they're likely to ultimately get infected by others anyway?

- Suppose an infected node infects a susceptible neighbor with independent probability β.
- Consider the ER network where there is a link from i to j iff i will infect j conditional on the event that i becomes infected while j is still susceptible.
- With any small number of initial seeds, the share of the population that ultimately gets infected will equal the size of the giant component of this network if at least one seed is in the giant component, and will equal 0 otherwise.
- Thus, the advantage of optimal seeding is that we can be sure to put at least one seed in the giant component, while with random seeding we might miss it...

Just a Few Seeds More? (cntd.)

- . . . However, if fraction q of the network is in the giant component, the probability of missing the giant component with s seeds is only $(1-q)^s$, which falls exponentially in s.
- With s random seeds, the expected ultimate reach of the diffusion is $1-(1-q)^s$ times as large as it is under optimal seeding.
- Hence, random seeding with just a few extra seeds does almost as well as optimal seeding.
- This suggests that optimal seeding may not be a big deal after all.

Caveats

It's interesting that random seeding does well in the standard diffusion model where each infected node infects a susceptible neighbor with independent probability β.

However, several caveats are in order before interpreting this as saying that optimal seeding doesn't matter in reality.

- Seeding does matter if individuals don't listen to each other with the same probabilities.
  - If everyone listens to the village elder and she doesn't listen to anyone, it's important to seed her.
- Seeding does matter under **complex contagion**, where you must be infected by several neighbors to become infected yourself (i.e., the number of infected neighbors must hit a **threshold** before you become infected).
  - Likely realistic especially in settings here agents are hesitant to adopt, like adopting an unproven, high-stakes technology.

Complex Contagion in Practice

Economists Beaman, BenYishay, Magruder, and Mobarak (2021) study the diffusion of a novel agricultural technique ("pit planting") among farmers in rural Malawai.

- They gather social network data and introduce two seeds (farmers trained in pit planting) in each village.
- In some villages, choose the seeds that maximize simple contagion (switch to pit planting if ≥ 1 neighbor uses it); in others, choose seeds that maximize complex contagion (switch iff ≥ 2 neighbors use pit planting).
- Find that adoption is significantly greater under the latter policy.
- Adoption rates are also closer to those predicted by complex contagion (lower than predicted by simple contagion).
- This suggests that something like complex contagion may be going on. It is also further evidence that seeding matters.

Diffusion Interventions: Summary

- Diffusions in societies and networks are often influenced by deliberate interventions to hinder or help their spread.
- Removal of high-degree nodes can greatly limit contagion, especially in networks with skewed degree distributions.
- Endogenous vigilance/social distancing helps explain the path of major epidemics like Covid-19.
- Targeted advertising or seeding is important for spreading information about new products or technologies, especially under complex contagion.

Morris Contagion Model

Another type of complex contagion is the diffusion of behaviors with a "coordination" aspect.

- E.g. want to use an operating system or social media platform that your friends use, want to consume the same products your friends consume.

In such settings, an individual usually wants to change her behavior only if **sufficiently many** of her neighbors change their behavior.

- This is sometimes called a **threshold model**.
  It is a close cousin of complex contagion.

This was formalized in a uniform-population model (without network structure) by sociologist Mark Granovetter in 1978.

- Similar ideas were introduced earlier by economist Thomas Schelling, whom we'll meet again later in the course.

The spread of such coordination behaviors on networks was introduced in an elegant model by Stephen Morris in 2000.

Contagion Model

Society represented by an undirected graph $G = (N, E)$.

- Each agent (node) $i \in N$ chooses an action $a \in \{0, 1\}$.
- Agent i wants to take action 0 iff less than fraction $q \in (0, 1)$ of her neighbors take action 1.
- She wants to take action 1 iff more than fraction q of neighbors take action 1.
- If exactly fraction q of neighbors take action 1, willing to take either action.

Note: in this model, q is a measure of the "inherent quality" of action 0, relative to action 1.

- $q < \frac{1}{2}$ means action 1 is "superior," in that it's optimal if one's neighbors are evenly divided;
  $q > \frac{1}{2}$ means action 0 is superior.

Contagion Model (cntd.)

What are the steady states (or **equilibria**) of this model?

- Clearly, everyone choosing 0 and everyone choosing 1 are both equilibria.
- Are there other equilibria, where some "groups" choose 0 and others choose 1?
- If we start at the 0 equilibrium (say) and some small group starts playing 1, do others switch?

Equilibria

- Recall that N_i is the set of i's neighbors.
- Say a set $S \subseteq N$ is **r-cohesive** if, for every $i \in S$, at least fraction r of i's neighbors are also in S: formally,

$$\min_{i \in S} \frac{|N_i(G) \cap S|}{|N_i(G)|} \geq r.$$

Theorem

For any set $S \neq \emptyset$, it is an equilibrium for everyone in S to play 1 and everyone else to play 0 if and only if S is q-cohesive and $N \backslash S$ is $(1-q)$-cohesive.

- Two requirements for equilibrium: no one in S wants to switch to $a = 0$, and no one in $N \backslash S$ wants to switch to $a = 1$.
- The first requirement is precisely that S is q-cohesive.
- The second requirement is precisely that $N \backslash S$ is $(1-q)$-cohesive.

Equilibria (cntd.)

Theorem

For any set $S \neq \emptyset$, it is an equilibrium for everyone in S to play 1 and everyone else to play 0 if and only if S is q-cohesive and $N \backslash S$ is $(1-q)$-cohesive.

Interpretation: The more highly cohesive sets there are in the network, the more scope there is for diverse patterns of behaviors.

- A network with many tight-knit communities can exhibit diverse behaviors, where some pockets use one technology/consume one type of product, and other pockets use different technologies/products.
- Networks with more uniform interaction structures tend to exhibit more homogeneous behaviors.

Contagion

Consider a variant where we start from the equilibrium where everyone plays $a = 0$ and we seed some set S with $a = 1$.

- Then, each individual with $a = 0$ switches to $a = 1$ iff more than fraction of q of its neighbors are playing $a = 1$.
- Repeat this process until no one wants to switch to $a = 1$.
- **Note:** we don't allow switching back from $a = 1$ to $a = 0$. This is sometimes called **progressive diffusion**.
  (To ponder: what changes in the alternative, non-progressive case?)
- We say there is **contagion from S** if this process results in $a = 1$ taking over the network.

Theorem

There is contagion from S iff, for any $S' \supseteq S$, the set $N \backslash S'$ is not $(1 - q)$-cohesive.

- Necessity is obvious: if for some $S' \supseteq S$, the set $N \backslash S'$ is $(1 - q)$-cohesive, contagion can never enter the set $N \backslash S'$.

Contagion (cntd.)

Theorem

There is contagion from S if and only if, for any $S' \supseteq S$, the set $N \backslash S'$ is not $(1-q)$-cohesive.

Proof of sufficiency:

- In every period, set of players taking $a = 1$ is some $S' \supseteq S$.
- If the set $N \backslash S'$ is not $(1-q)$-cohesive, some player will switch to $a = 1$.
- Since the network is finite, eventually everyone will have switched to $a = 1$.

This is a powerful theorem. However, in practice it can be difficult to check whether all subsets of N_{26} are q- or $(1-q)$-cohesive.

Remark

Recall that low q means action 1 is "superior."

- Theorem says that even superior innovations may not spread.
- However, the better the innovation (the lower is q), the more cohesive the network must be to stop it from spreading.
  - E.g. in the complete network with n nodes, if infect m initial nodes, the complementary set of $n-m$ agents is $\frac{n-m-1}{n-1}$ cohesive. So the innovation spreads iff $q < \frac{m}{n-1}$.
- Conversely, inferior innovations can spread.
- However, we will see a limit to this result on the next slide.

Infinite Graphs

We can apply the same contagion process to an infinite graph (perhaps a better model for contagion in a large society than a finite graph).

- In this case, by **contagion from** S we mean that given the (finite) initially infected set S, the infection grows without bound.
- E.g. if the graph is an infinite line and $q = \frac{1}{2} - \varepsilon$, contagion spreads from a single initially infected node.

However, we can obtain the striking result that an inferior innovation (that is, one with $q > \frac{1}{2}$) can **never** spread.

Infinite Graphs

We can apply the same contagion process to an infinite graph (perhaps a better model for contagion in a large society than a finite graph).

- In this case, by **contagion from** S we mean that given the (finite) initially infected set S, the infection grows without bound.
- E.g. if the graph is an infinite line and $q = \frac{1}{2} - \varepsilon$, contagion spreads from a single initially infected node.

However, we can obtain the striking result that an inferior innovation (that is, one with $q > \frac{1}{2}$) can **never** spread.

Theorem

For any infinite graph where each node has finite degree, if $q > \frac{1}{2}$ *then, for any finite initially infected set* S*, contagion does not spread (infinitely far) from* S*.*

Infinite Graphs (cntd.)

Proof.

At any period, define the **interface** between the infected and non-infected region to be the set of all *links* between an infected node and a non-infected node.

Claim: in each period, either no new nodes become infected or the number of links in the interface strictly decreases.

- Since $q > \frac{1}{2}$, if any new node becomes infected in period $t+1$, it must have strictly more links to the period-t infected set than to the period-t non-infected set.
- The former links were in the period t interface but not the period $t+1$ interface. The latter links were not in the period t interface and may or may not join the interface in period $t+1$. No other links join the interface.

Since the number of links in the period 0 interface is finite, there can only be finitely many periods at which new nodes become infected.

Morris Contagion Model: Summary

- The Morris contagion model is a model of the network diffusion of behaviors with a coordination aspect.
- The more cohesive is the network, the more likely there are to be multiple equilibria with different patterns of behavior.
- The more cohesive is the network, the harder it is for a new innovation to spread, even if it is superior to the status quo.
- For some networks, an inferior innovation can spread if the set of initial adopters is large enough.
- However, an inferior innovation can never spread infinitely far.

14.15 / 6.207 Networks

Spring 2022