Network Dynamics I: Bayesian Learning, Information Cascades

Web Science (VU) (706.716)

Elisabeth Lex

ISDS, TU Graz

May 14, 2018

Repetition

- Random Graphs
- Small World Phenomenon
- Power Laws & Preferential Attachment

Motivation and Introduction

Motivation

- Individuals easily influenced by decisions of others, especially in social and economic situations.
- Ex.: opinions, products to buy, political positions

Motivation

- Individuals easily influenced by decisions of others, especially in social and economic situations.
- Ex.: opinions, products to buy, political positions
- Today: Why does such influence occur and how can we model this?

Example: Look Up

- In the 60s, some experimenters had groups of people 1-15 look up in the sky.
- They then watched to see if anyone else looked up as well.

Example: Look Up

- In the 60s, some experimenters had groups of people 1-15 look up in the sky.
- They then watched to see if anyone else looked up as well.
- With one looking up, most ignored him.

Example: Look Up

- In the 60s, some experimenters had groups of people 1-15 look up in the sky.
- They then watched to see if anyone else looked up as well.
- With one looking up, most ignored him.
- With 5 looking up, many stopped.

Example: Look Up

- In the 60s, some experimenters had groups of people 1-15 look up in the sky.
- They then watched to see if anyone else looked up as well.
- With one looking up, most ignored him.
- With 5 looking up, many stopped.
- With 15, 45% stopped and kept looking up.

Example: Restaurant

- You go to a unfamiliar city such as San Francisco.
- Where do you decide to eat?
- Restaurant A and B both look nice on the outside and have a similar menu.
- A has a small crowd.
- B is empty.
- Which do you go to?

Example: Restaurant

- You go to a unfamiliar city such as San Francisco.
- Where do you decide to eat?
- Restaurant A and B both look nice on the outside and have a similar menu.
- A has a small crowd.
- B is empty.
- Which do you go to?
- Probably A.

Example: Restaurant

- You go to a unfamiliar city such as San Francisco.
- Where do you decide to eat?
- Restaurant A and B both look nice on the outside and have a similar menu.
- A has a small crowd.
- B is empty.
- Which do you go to?
- Probably A.
- But: What if from the outside B looks slightly better?
- How about if some clerk at the hotel said he heard B was good?

Observations

- In all these cases, decisions are made sequentially.
- People make decisions based on inferences from what earlier people have done
- Individuals may imitate behavior of others but not mindless

Observations

- In all these cases, decisions are made sequentially.
- People make decisions based on inferences from what earlier people have done
- Individuals may imitate behavior of others but not mindless
- Sometimes it is rational for an individual to follow the crowd even if the individual's own information suggests an alternative choice

Reasons for Imitation

- Direct-Benefit Effects: Actions of others affect you *directly*
- E.g. Becoming part of a social networking site - choosing an option that has a large user population
- Informational Effects: Actions of others affect you *indirectly* by changing your information

How to model this?

- Decision-based Models: adopt new behaviors if k others do it
- Probabilistic Models: adopt a behavior ("catch a disease") with some probability from neighbors in the network

Herding & Information Cascades

Herding

- A **decision** needs to be made
- People make the decision **sequentially**
- Each person has some **private information** that helps with the decision
- This private information **cannot be directly observed** but one can see what people **do**
- This way, inferences about their private information can be made

A Simple Herding Example

Large group of students - participants

- Consider two urns with 3 balls:
  - Majority-blue: 2 blue 1 red
  - Majority-red: 2 red, 1 blue
- Each person wants to best guess whether the urn is majority-blue or majority-red:
- Experiment: One by one each person
  - Draws a ball
  - Privately looks at its color and puts it back
  - Publicly announces her guess
- Students receive monetary rewards for correct guesses

What happens? (1/2)

- 1st person: guess the color drawn
- 2nd person: guess the color drawn
- 3rd person:
  - If the two before made different guesses, then go with her own color
  - Else: just go with their guess (regardless of the color you see) - student can reason that the urn is majority-blue

What happens? (1/2)

- 1st person: guess the color drawn
- 2nd person: guess the color drawn
- 3rd person:
  - If the two before made different guesses, then go with her own color
  - Else: just go with their guess (regardless of the color you see) - student can reason that the urn is majority-blue
- In the latter case, an *information cascade* has begun: the 3rd person makes same guess as the first 2, regardless her own private information

Information Cascades

Definition

An information cascade develops when people abandon their own information in favor of inferences based on earlier people's actions

What happens? (2/2)

4th person and onward:

- Say first 2 guesses were both blue, 3rd person will also guess blue, regardless of what she saw
- 4th heard blue 3 times in a row, but knows that 3rd guess conveys no information
- 4th in the same situation as 3rd - should also guess blue
- Continues with all subsequent students since everyone's best strategy is to rely on limited genuine information available

Modeling this type of reasoning: Bayes's Rule (1/2)

- We want to build a mathematical model to how information cascades occur
- We need to compute probabilities of events (e.g. event is "the urn is majority-blue")
- Whether an event (not) occurs is the result of certain random outcomes (e.g. which urn was placed in the room)
- We image a large sample space: each point in this space is a realization for each of the random outcomes

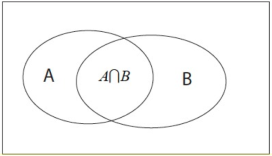
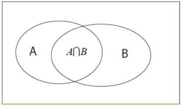


Figure: Two events A and B in a sample space. Region of A = set of all outcomes where A occurs. $A \cap B$ is where both A and B occur

Modeling this type of reasoning: Bayes's Rule (2/2)

We need to estimate the **conditional probability** of event A given that event B has occurred:



- The fraction of the area of region B occupied by the joint event $A \cap B$:

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \tag{1}$$

- We can follow the Bayes' rule:

$$P[A|B] = \frac{P[A] * P[B|A]}{P[B]} \tag{2}$$

where $P[A]$ is the prior probability of A and $P[A|B]$ is the posterior probability of A given B

Our Example with Bayes's Rule (1/3)

Each person tries to estimate **conditional probability** that urn is majority-blue or majority-red given what she has seen and heard

She should guess

- P_r[majority-blue|what she has seen and heard] $> \frac{1}{2}$]
- Majority-red otherwise

If both conditional probabilities 0.5 - doesn't matter what she guesses

Our Example with Bayes' Rule (2/3)

- Prior probabilities $P_r[\text{majority-blue}]$ and $P_r[\text{majority-red}] = \frac{1}{2}$
- Probabilities for the 2 urns $P_r[\text{blue}|\text{majority-blue}] = P_r[\text{red}|\text{majority-red}] = \frac{2}{3}$

Let's assume that 1st person draws blue ball

- $P_r[\text{majority-blue}|\text{blue}] = \frac{P_r[\text{majority-blue}] * P_r[\text{blue}|\text{majority-blue}]}{P_r[\text{blue}]} = \frac{2}{3}$

Ergo: Conditional probability greater than $\frac{1}{2}$, 1st person should guess "blue"

Our Example with Bayes' Rule (3/3

Calculation for 3rd person:

- $P_r[\text{majority-blue}|\text{blue, blue, red}] = \frac{P_r[\text{majority-blue}] * P_r[\text{blue, blue, red}|\text{majority-blue}]}{P_r[\text{blue, blue, red}]} = \frac{2}{3}$

Ergo: Conditional probability greater than $\frac{1}{2}$, 3rd person should guess "blue"

Discussion

- Cascade can occur easily given the right structural conditions
- Can lead to non-optimal outcomes
- In our example, with prob $1/3x1/3 = 1/9$, the first two would see the wrong color, from then on, the whole population would guess wrong
- Can be very fragile despite their potential to produce long runs of conformity
  - Suppose, first 2 guesses are blue
  - People 100 and 101 draw red and cheat, i.e. they show the drawn balls
  - Person 102 has then 4 pieces of honest information - she should guess based on her own color
  - Cascade is broken!

A general formulation of the model

A General Cascade Model

- Group of people sequentially make decisions
- Decision is rejecting or accepting some option
- **First Model Ingredient:** States of the World. A world is placed into one of two states: if the option is actually a good idea G (with probability p), or a bad idea B (with probability $1 - p$)
- **Second Model Ingredient:** Payoffs. Each individual receives a payoff based on her decision. If she chooses to reject, payoff is 0. If she chooses to accept in case of G, the payoff obtained from accepting it is a positive number $v_g > 0$. If she chooses to accept while B, the payoff is $v_b < 0$

A Simple General Cascade Model

- **Third Model Ingredient:** Signals. To model the private information, 2 possible signals:
- High signal (H) - suggests accepting is a good idea
- Low signal (L) - suggests accepting is a bad idea
- If accepting is a good idea: high signals more frequent, i.e., $P[H|G] = q > 1/2$, $P[L|G] = 1 - q < 1/2$
- If accepting is a bad idea, low signals more frequent, i.e., $Pr[L|B] = q$ and $Pr[H|B] = 1 - q, q > 1/2$

Link to Herding Experiment

- Two possible states of the world: urn is majority-blue or majority-red
- Accepting corresponds to guessing majority-blue
- State G (good idea) if urn is majority-blue
- State B (bad idea) otherwise
- Prior probability of accepting being a good idea is $p = 1/2$
- Private information in the experiment is the color of the ball the individual draws:
- H if signal is blue, thus: $P[H|G] = P[blue|majority - blue] = q = 2/3$

Modeling the Individual Decisions

- Suppose a person gets signal H
- Payoff: $P[G|H] + P[B|H]$
- Calculation: $P[G|H] = \frac{P[G]*P[H|G]}{P[H]}$
- Denominator: $P[H] = P[G]*P[H|G] + P[B]*P[H|B]$
- If we substitute $P[H|G]$ with q, that gives us : $\frac{pq}{pq+(1-p)(1-q)}$:
- Final inequality: $pq + (1-p)(1-q) < pq + (1-p)q = q$

Intuition: high signal more likely to occur if option is good. If an individual observes a high signal, they raise their estimate of the probability that the option is good. Expected payoff shifts from 0 to positive number - they should accept the option

Multiple Signals

Let us assume sequence S of independently generated signals consisting of a high signals and b low signals, interleaved in some fashion

- Posterior probability $P[G|S]$ is greater than prior $P[G]$ when $a > b$
- Posterior $P[G|S]$ is less than prior $P[G]$ when $a < b$
- Probabilities $P[G|S]$ and $P[G]$ equal when $a = b$

Multiple Signals

Let us assume sequence S of independently generated signals consisting of a high signals and b low signals, interleaved in some fashion

- Posterior probability $P[G|S]$ is greater than prior $P[G]$ when $a > b$
- Posterior $P[G|S]$ is less than prior $P[G]$ when $a < b$
- Probabilities $P[G|S]$ and $P[G]$ equal when $a = b$

Individuals should accept the option if the get more high signals than low ones and reject it when they get more low signals than high signals. They are indifferent when they get the same number of highs and lows.

Prove that with Bayes' Rule

- $P[G|S] = P[G] * P[S|G]/P[S]$
- Nominator: Prior of G multiplied by $P[S|G]$
- since signals are generated independently, we can simply multiply their probabilities: this gives us a factors of q and b factors of $(1-q)$
- Thus: $P[S|G] = q^a(1-q)^b$
- $P[S] = P[G] * P[S|G] + P[B] * P[S|B]$ to cover that S can come up in both cases G and H, which gives: $pq^a(1-q)^b + (1-p)(1-q)^a q^b$
- $P[G|S] = pq^a(1-q)b/pq^a(1-q)^b + (1-p)(1-q)^a q^b$

Evaluation

We want to know how $P[G|S] = pq^a(1-q)b/pq^a(1-q)^b + (1-p)(1-q)^a q^b$ compares to p and q

- Approach: replace second term in the denominator by $(1-p)q^a(1-q)b$
- This gives: $pq^a(1-q)^b + (1-p)q^a(1-q)^b = q^a(1-q)^b$
- Replacing this in the equation $P[G|S] = pq^a(1-q)b/q^a(1-q)^b = p$

Evaluation

We want to know how $P[G|S] = pq^a(1-q)b/pq^a(1-q)^b + (1-p)(1-q)^a q^b$ compares to p and q

- Approach: replace second term in the denominator by $(1-p)q^a(1-q)b$
- This gives: $pq^a(1-q)^b + (1-p)q^a(1-q)^b = q^a(1-q)^b$
- Replacing this in the equation $P[G|S] = pq^a(1-q)b/q^a(1-q)^b = p$
- If $a > b$: denominator becomes larger, overall expression smaller, hence $P[G|S] > p = P[G]$, if $a < b$ otherwise and if $a = b$, the expression evaluates to p

Sequential Decision-Making and Cascades

Model how individuals should make decisions about rejecting & accepting in sequence

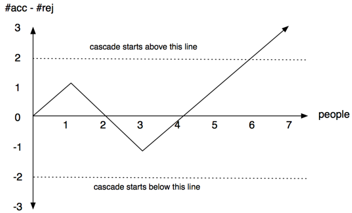
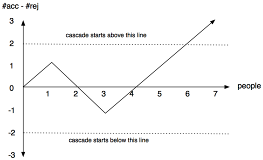


Figure: Cascade begins when diff between nr of accepts and rejects reaches 2.

- P 1 will follow his own private signal
- If signal P 2 receives differes from P 1's, person 2 follows her own
- P 3 will either follow the majority signal or if P 1 and P 2 made opposite decisions, P 3 will use own signal as tie-breaker.

Sequential Decision-Making and Cascades

Let's consider the perspective of a person numbered N



- If nr of accepts among people before N is equal to nr of rejects, then N's signal is tie-breaker - N follows her own signal
- If nr of accepts among people before N differs from nr of rejects by 1, N follows own signal
- If nr of accepts among people before Ndiffers from number of rejects by 2 or more, then N's signal won't outweigh this earlier majority. N follow the majority
- Then, $N+1, N+2, ..$ know that N ignored her own signal so they'll follow the majority, and hence a cascade has begun.

Summary

- As long as nr of acceptances differs from nr of rejections by at most one, each person in sequence follows their own private
- Once nr of acceptances differs from nr of rejections by two or more, a cascade takes over, everyone follows the majority decision
- When people can see what others do but not what they know, initially, they rely on their own private information.
- Over time, population begins ignoring own information and following the crowd - while still being fully rational

Summary

- Cascades can be wrong
- Cascades can be based on very little information
- Cascades are fragile
- One has to be careful in drawing conclusions about the best course of action from the behavior of a crowd. As we have just seen, the crowd can be wrong even if everyone is rational and everyone takes the same action.

Applications and Practical Value

- Early adopters in Online Marketing: attempting to initiate a buying cascade
- Collaboration and consensus building: can be wise to make collaborators reach partial conclusions before entering phase of collaboration (e.g. hiring committees)

Extensions of Basic Cascade Model: Threshold Models

How should we organize revolt?

- You live an in oppressive society
- You know of a demonstration against the government planned tomorrow
- If a lot of people show up, the government will fall
- If only a few people show up, the demonstrators will be arrested and it would have been better had everyone stayed at home

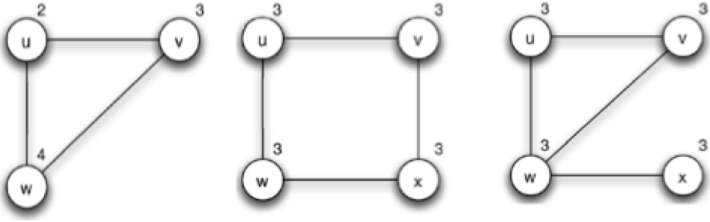
Pluralistic Ignorance

- You should do something if you believe you are in the majority!
- Pluralistic ignorance – erroneous estimates about the prevalence of certain opinions in the population
- Survey conducted in the U.S. in 1970 showed that while a clear minority of white Americans at that point favored racial segregation, significantly more than 50% believed it was favored by a majority of white Americans in their region of the country.

Organizing the revolt: A Model

- Personal threshold k: "I will show up if am sure at least k people in total (including myself) will show up"
- Each node only knows the thresholds and attitudes of all their direct friends.
- Can we predict if a revolt can happened based on the network structure?

Which Network Can Have a Revolt?



Linear Threshold Model (1/2)

One of the most popular diffusion models

- Actor takes action if nr of neighbors (friends), who also take the action is greater than a certain threshold
  - Each directed edge $(u, w) \in E$ has non negative weight $b(u, w)$
  - For each node $v \in E$, total incoming edge weights sum to less than or equal to one
  - Each node v chooses threshold t_v randomly from a uniform distribution in an interval between 0 and 1
  - Then, in each step, all nodes remain active that were active in the previous step
  - Nodes that satisfy the following condition will be activated: $\sum_{w \in N_v, \text{w is active}} b_{w,v} \geq t_v$

Linear Threshold Model (2/2)

- An inactive node is influenced by all of its active neighbors at each time step
- An active node influences its inactive neighbors according to the weights
- At each step, an inactive node becomes active if the total weight of its incoming neighbors is at least t_v
- Thresholds t_v selected randomly due to lack of knowledge of the tendency of node
- Express different levels of tendency of nodes to adopt an idea or innovation

Linear Threshold Model: Example

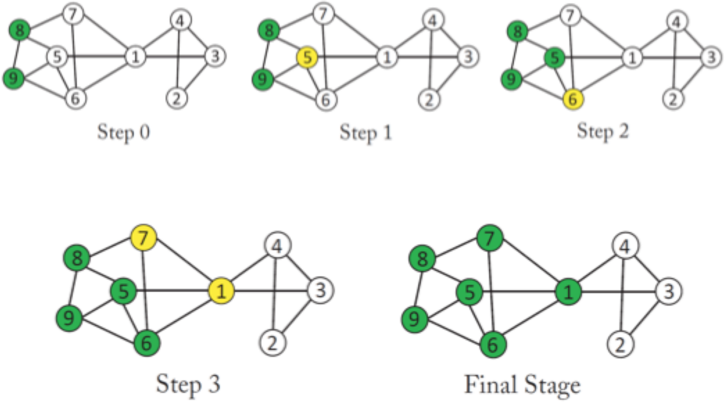


Figure: Linear Threshold Model Diffusion Process

Probability based models

Probabilistic models

- Models of influence or disease spreading
- An infected node tries to “push” the contagion to an uninfected node
- Ex.: Catching a disease with some probability from each active neighbor in the network

Probabilistic Models

- Aka Contagion, Propagation: Randomly occur as a result of social contact - no decision making involved
  - SIS: Susceptible-Infective-Susceptible (e.g., flu)
  - SIR: Susceptible-Infective-Recovered (e.g., chickenpox)
- Question: Will the virus take over the network?
- Independent contagion model

Propagation

- How do viruses spread?
- How do rumors propagate?
- Will a flu-like virus linger or will it die out soon?
- Birth rate β: probability that infected neighbor attacks
- Death rate δ: probability that infected node heals

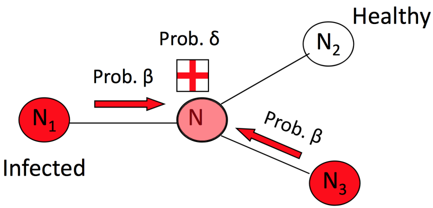


Figure: Modeling propagation

SIR: Susceptible-Infective-Recovered Model

SIR is suitable for modeling a disease that each individual can only catches once during their life time.

- Initially, some nodes are in the I state and all others in the S state.
- Each node v in the I state remains infectious for a fixed number of steps t
- During each of the t steps, node v can infect each of its susceptible neighbors with probability p.
- After t steps, v is no longer infectious or susceptible to further infections and enters state R.

SIS: Susceptible-Infective-Susceptible Model

- Each node may be healthy (susceptible) or infected
- Cured nodes immediately become susceptible
- Virus strength: $s = \beta/\delta$ (spreading rate)
- Epidemic threshold τ: if $s = \beta/\delta < \tau$, epidemic cannot happen

Epidemic Threshold

What should τ be dependent on?

- Average degree? Highest degree?
- Variance of degree?
- Diameter?

Epidemic Threshold

Theorem by Wang et al 2003: No epidemic if

$$\frac{\beta}{\delta} < \tau = \frac{1}{\lambda_{1,A}} \quad (3)$$

where $\lambda_{1,A}$ is the largest eigenvalue of the adjacency matrix A

Real-World Application of Diffusion Processes: Viral Marketing

Viral Marketing and Recommender Systems

How do recommendations and purchases propagate? [Leskovec, Adamic, Huberman 2006]

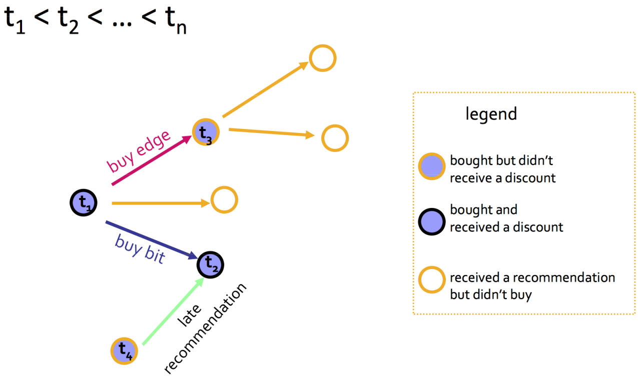
- Product recommendation network
- Viral Marketing[1]
- Senders and followers of recommendations receive discounts on products
- Recommendations are made to any number of people at the time of purchase
- Only the recipient who buys first gets a discount

[1] The dynamics of viral marketing, J. Leskovec

Dataset

- Large anonymous online retailer (June 2001 to May 2003)
- 15,646,121 recommendations
- 3,943,084 distinct customers
- 548,523 products recommended
- 4 product groups: books, DVDs, music, VHS

Identifying the cascades



Role of product category

	products	customers	recommenda-tions	edges	buy + get discount	buy + no discount
Book	103,161	**2,863,977**	5,741,611	2,097,809	**65,344**	**17,769**
DVD	**19,829**	805,285	**8,180,393**	962,341	**17,232**	**58,189**
Music	**393,598**	794,148	1,443,847	585,738	7,837	2,739
Video	26,131	239,583	280,270	160,683	909	467
Full	542,719	3,943,084	15,646,121	3,153,676	91,322	79,164

Figure: Red is high, blue is low. People are nodes, edges are recommendations

Recommendations for DVDs are more likely to result in a purchase

Some observations

- Few DVDs, but DVDs make approx. 50% of recommendations
- Music recommendations reached approx. same nr of people as DVDs but used only 1/5 as many recommendations
- Book recommendations reached the most people
- Networks highly disconnected
- Overall, 3.69 recommendations per node on 3.85 different products

Viral Marketing

Does sending more recommendations influence more purchases?

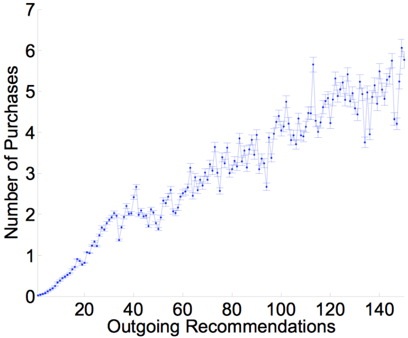


Figure: Influence of number of sent recommendations on purchases

Summary

- Herding and Information Cascades, Bayes Rule as Model
- Decision based models: Threshold Models, Linear Threshold Model
- Probability based models: SIR, SIS
- Viral Marketing as application scenario

Thanks for your attention - Questions?

elisabeth.lex@tugraz.at

Slides use figures from Chapter 16 and Chapter 19 of Networks, Crowds and Markets by Easley and Kleinberg (2010)
http://www.cs.cornell.edu/home/kleinberg/networks-book/
http://web.stanford.edu/class/cs224w/handouts.html
http://snap.stanford.edu/na09/11-viral-annot.pdf