Social Influence

- Topology-Dependent Static Influence Models
- Link Importance: Strong and Weak Ties
- Influence Dynamics in Social Networks and Applications:
- Epidemic-Based Models
- Activation-Based Models (Discrete Time)
- Models for Opinion Dynamics
- Continuous-Time Diffusion Models
- Influence Maximization
- Influence: Theory versus Practice

Influence in Social Networks

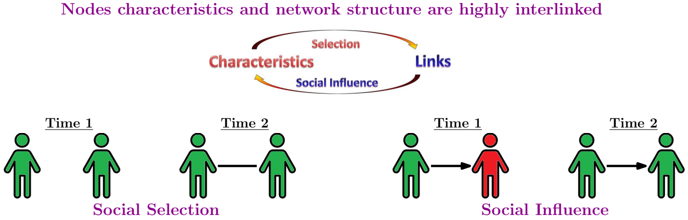
Social Network Influence

Social network influence is the process by which individuals or entities affect the behaviors, opinions, or actions of others within a network.

Influence can be direct or indirect, where the influence spreads through connections or "ties" between people or nodes

Two important phenomena in Sociology

- Social Selection: Individual's attributes drive the interaction with others
- Social Influence: Interactions among people shape people's attributes



Importance of Social Network Influence

Social network influence plays a crucial role in various domains. Understanding the mechanics of influence can help in optimizing strategies for widespread adoption of ideas, products, or behaviors.

Influence is both a social and algorithmic phenomenon, involving key concepts like centrality and content virality

- Marketing Campaigns: Influential users can amplify companies' products
- Political Movements: Influential political leaders use social networks to sway opinions mobilize support, and encourage people to participate in events, protests, or movements
- Health Information Spread: Public health officials leverage influencers to spread health tips or warning messages (e.g., vaccination campaigns)
- Content dissemination: Influencers (celebrities or thought leaders) help the spread of information/misinformation
- Behavioral Change: Users tend to mimic influencers' behaviors. Product promoted by an influencer may prompt their followers adoption

Influence Diffusion Process

The study of influence in social networks focuses on how individuals, groups, or organizations affect each other's behavior, opinions, or decisions

Influence can spread across various connections in a network, often driven by social interactions, peer pressure, or other factors

Understanding influence diffusion process in social networks helps us answer important questions about the spread of information, behaviors, and beliefs

- Predict the spread of behaviors: Understand how behaviors like voting, disease, or adoption of new technologies spread.
- Optimize marketing strategies: By understanding how influence spreads, companies can target the right individuals to promote products
- Maximize influence: Determine which individuals (seed nodes) to target to maximize the spread of influence in a network
- Prevent the spread of misinformation: Influence models help identify key nodes that could prevent the rapid spread of rumors or false information.
- Understand social dynamics: These models help explain how behaviors or opinions change and spread within communities.

Key Questions We Want to Answer with Influence Diffusion Models

Influence diffusion models help answer key questions:

- How can we maximize influence in a network?
- Which individuals in a network are most influential?
- How does influence dissipate over time and distance in a network?
- How can we prevent the spread of negative or harmful information?
- What are the effects of network structure on the diffusion of influence?

These questions are crucial for both business applications and social science research

Where Do These Models Fit in Social Network Analysis?

- Network Dynamics: These models show how networks evolve over time as influence spreads among nodes
- Community Behavior: Influence models help explain how communities form, interact, and influence each other
- Resource Allocation: These models are applied in scenarios where resources need to be spread across a network (e.g., information, vaccines, marketing content)
- Opinion Formation: Understanding how individuals' opinions evolve based on their interactions with neighbors in the network

By studying these models, researchers can predict outcomes and design better interventions for social and organizational networks

Traditional Media Influence vs. Social Media Influence

Traditional Media Influence:

- Centralized Control on content distribution in traditional media (TV, radio, newspapers)
- Top-Down Communication from media outlets to audience
- Gatekeeping: Editors and producers control, filter and censor content via editorial policies– credible
- Mass Reach but Limited Interaction between content producer and (generally) large audience
- Slow Speed of Information Dissemination

Social Media Influence:

- Decentralized Communication: Users generate and share content freely, reducing the control of gatekeepers, variable credibility
- Two-Way Communication between content creators and consumers, facilitating discussions and reactions in real-time
- Viral Potential of information on social media through shares, reposts, likes, and retweets
- Rapid Spread of User-Generated Content via engagement

Influence Analysis Problems

In social network analysis, there are two main types of influence problems: the analysis problem and the synthesis problem. These two areas help us understand both the measurement and application of influence in networks.

- Analysis Problem: How do we measure or quantify the influence within a network?
- Synthesis Problem: How can we use the knowledge of influence to actively influence or control the network?

Key questions in Social Network Influence:

- How do we quantify influence in social networks?
- What are the mathematical models used to understand influence?
- How does network structure affect the spread of influence?

Analysis Problem: Measuring Influence

The analysis problem focuses on measuring influence in a network. Some common ways to approach this problem include:

- Node Centrality: Identifying which nodes are the most influential in the network
- Link Strength: Measuring the importance of the connections between nodes
- Diffusion Patterns: Analyzing how information or behaviors spread through the network

Key tools for measuring influence:

- Graph Theory: Used to model networks and quantify influence using various centrality measures
- Statistical Models: Used for understanding patterns of influence over time

Synthesis Problem: Leveraging Influence

The synthesis problem involves using the understanding of influence to actively control or influence a network

Key aspects of this problem include:

- Targeted Influence: Choosing which nodes to influence in order to maximize the spread of information or behavior
- Influence Maximization: Identifying the optimal set of nodes to influence in a given budget or time constraint
- Viral Marketing: Using influential nodes to initiate a viral spread of content

Approaches to solving the synthesis problem:

- Optimization Algorithms: Used to maximize influence over a network with limited resources
- Heuristic Methods: Approaches based on empirical observations and trial strategies

Theory vs. Practice in Influence Models

There is a significant gap between the theoretical models of influence and their practical implementation in real-world networks

Mathematical models provide a framework for understanding the spread of influence, their direct application often requires simplifications, approximations, and empirical adjustments

Key challenges in bridging theory and practice:

- Complex Network Structures: Real-world networks often have complex and dynamic structures that theoretical models may not fully capture
- Data Limitations: Access to complete data on network topologies and user behaviors is often limited, making real-world modeling difficult
- Dynamic Behavior: Users in social networks often change behaviors over time, which is difficult to model with static theories

Despite these challenges, the principles behind influence models are widely used in practice, particularly in marketing and social media campaigns

Topology-Based Static Influence Modeling

Topology-Dependent Static Influence Models

Topology-dependent influence models analyze how the structure (or topology) of a network affects the spread of influence across it

The nodes and edges of the network, and their arrangement, play a key role in determining how information, behaviors, or trends propagate

Key Idea: Influence in a network is determined not only by individual nodes but also by the overall network structure, such as the presence of hubs, bridges, and community clusters

Social media platforms, biological networks, and communication networks use these models to understand how the shape of the network influences behaviors like information diffusion, adoption of technologies, or the spread of epidemics

Static graph metrics (centrality measures) evaluate the influence of nodes based on the network's fixed structure

They are useful for evaluating influence in stable, unchanging networks

Direct Measures: Followers, Retweets, Reposts

Direct metrics are often used in marketing and social media analytics to quantify influence, although they may not always capture the full scope of a user's impact

- Followers: The number of people following a user on Twitter or Facebook
  - Influence Through Reach: More followers $\implies$ more visibility of content
  - Authority: Large followings often indicate that a user is seen as an authority or a trusted source in a specific domain
  - Engagement vs. Reach: Reach doesn't necessarily translate to actual engagement or trust or real influence
  - Fake Followers: Inflated follower count by buying fake followers
  - Contextual Influence: Doesn't account for influence in niche communities
- Retweets and Reposts: Frequency of a user's content being shared by others
  - Amplification Effect: Higher number of retweets or reposts suggests content is being amplified by other users, extending the original user's influence
  - Network Cascade: A single repost can trigger a cascade where followers of the reposting user share the content, leading to exponential visibility growth
  - Virality Potential: Posts that are shared frequently can go viral, allowing the content to reach audiences far beyond the original user's direct followers
- Mentions: How often a user's content is referenced or discussed by others

Node Centrality in Networks as Influence Measures

Node centrality (aka prestige in digraphs (Twitter/Web)) - structural importance of a user in a network as a proxy for potential influence

- Importance (functional) role of network players is often related to their (structural) position in the network
- It can significantly differ from presumption about importance e.g. fathers, mothers, executives, teachers, ...
- Centrality analytics undertakes quantitative social network analysis to determine types of actors and find key players

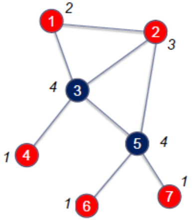
Structural and functional (dynamic) importance in essence are $f : V \mapsto \mathbb{R}$

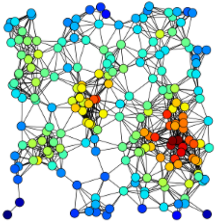
- Can be used to identify influential actors
- Robustness and vulnerability of network
- Determine exposure of nodes to disease or their role in immunization
- Study of spread and countering epidemic

Node Centrality in Networks as Influence Measures

Degree centrality: How many nodes can this node reach directly?

$$C_d(v) \; := \; deg(v)$$





- In a digraph we often use in-degree
- In Twitter or Web graph amounts to nodes popularity or influence
- Useful to determine important nodes for spreading information and influencing others in their immediate neighborhood

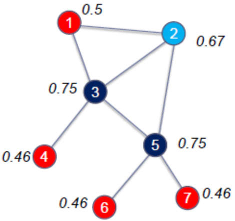
Node Centrality in Networks as Influence Measures

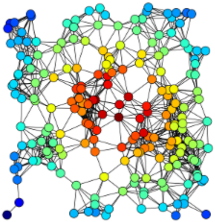
Closeness centrality: How fast can this node reach other nodes?

$$C_{close}(v) \;:=\; \frac{1}{\sum_u d(v,u)} \qquad \text{or} \qquad C_{close}(v) \;:=\; \frac{|V|}{\sum_u d(v,u)}$$

Assuming communication happens via shortest paths only, high closeness centrality nodes can reach other nodes the easiest- A measure of reach

To compare graph of varying orders one usually normalize





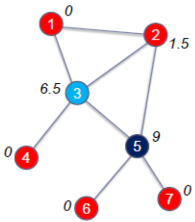
Node Centrality in Networks as Influence Measures

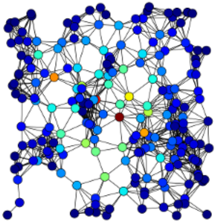
Betweenness centrality: likelihood of node to be on communication path

$$C_{bw}(v) := \sum_{s,t \neq v} \frac{\lambda_{st}(v)}{\lambda_{st}}$$

- $\lambda_{st}(v)$: Number of shortest path between s and t via v
- λ_{st}: Number of shortest path between s and t

Assuming communication happens via shortest paths only, high betweenness centrality nodes are critical for information flow





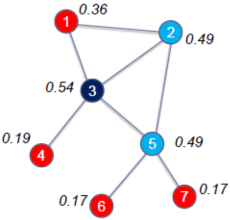
Node Centrality in Networks as Influence Measures

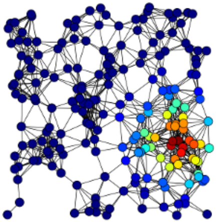
Eigenvector centrality: the node's connectivity to "well-connected" nodes?

Proportional to sum of eigencentralities of neighbors

$$\mathbf{c}(v) := \frac{1}{\lambda} \sum_{u \in N(v)} \mathbf{c}(u)$$

- λ is a constant ▷ leading eigen value
- Computed as $A\mathbf{c} = \lambda\mathbf{c}$ ▷ A is the adjacency matrix





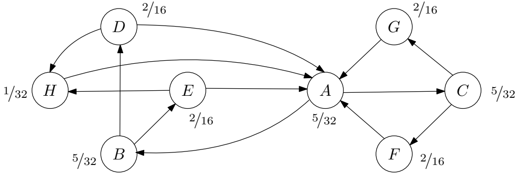
Node Centrality in Networks as Influence Measures

Pagerank centrality: the node's connectivity to "well-connected" nodes?

Proportional to weighted sum of pagerank of out-neighbors

$$\mathbf{c}(v) \; := \; \alpha \sum_{u \in N^-(v)} \frac{\mathbf{c}(u)}{deg^+(u)} + \frac{1-\alpha}{|V|}$$

- α is the damping factor
- Probability of a random walker to visit the node



Node Centrality in Networks as Influence Measures

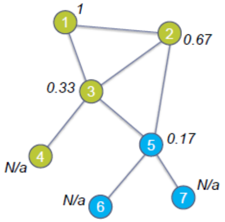
Clustering Coefficient: which nodes in the graph tend to cluster together

$$C(v) \;=\; \frac{|E(G[N(v)])|}{\binom{d(v)}{2}}$$

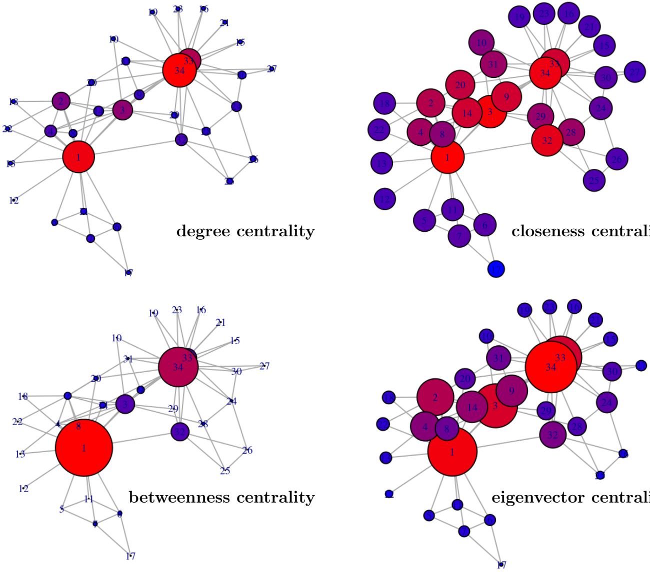
- $E(G[N(v)])$ is edges in graph induced by $N(v)$

number of triangles around a node v (friendships b/w v's friends)

Closely related to **transitivity** of a graph - ratio of observed number of closed triangles/triplets and max possible number of closed triplets



Node Centrality in Networks as Influence Measures



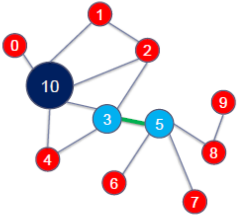
source: D. Petrov, Y. Dodonova, A. Shestakov (2015)

Caveats about Centrailities

- No single "right" centrality measure, each gives a different perspective
- Each centrality measure is a proxy of an underlying network process
- Unrealistic or irrelevant process lead to unrealistic centrality
- Centrality is used as graph EDA to gain insights about structure
- "Enhanced metrics" exist for graphs with more "features" (e.g. directed, weighted edges)
- Notion of centralities can be extended to edges

Identifying sets of key players

Importance of individual nodes may not reveal much



node 10 is most central, but node 3 and 5 together are more critical for network connectivity (the edge $(3, 5)$ is a bridge)

Medici Family's Influence in Renaissance Florence

15th Century Florentine Marriages

(Padgett & Ansell, 1993), cf. (Jackson, 2010)

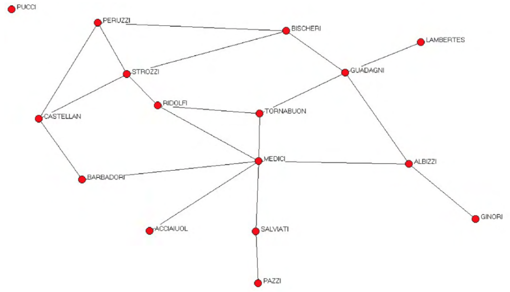


Figure 1.2.1 15th Century Florentine Marriges Data from Padgett and Ansell [491] (drawn using UCINET)

It illustrates the relationships between different families in Renaissance Florence, with nodes representing individual families and edges representing marriages between them

The families represented here were part of the political and economic elite of Florence, and their intermarriages helped consolidate power and influence. The Medici family, for example, used strategic marriages to build and maintain their power, which can be seen through their central position in the network.

Medici Family's Influence in Renaissance Florence

- The Medici family had the highest degree centrality, meaning they had the most direct connections
- Their strategic position as intermediaries between influential families gave them high betweenness centrality
- The Medici's proximity to other powerful families gave them significant closeness centrality

Limitations of Static Influence Models

While static models are useful for understanding influence at a single point in time, they have several limitations:

- Time Invariance: Static models do not capture changes over time. Influence can fluctuate as network structures evolve, making static metrics less useful for dynamic networks
- Over-simplification: Real-world networks often experience frequent changes, such as the creation and dissolution of connections. Static models ignore these temporal aspects
- Missed Cascades: Influence that spreads slowly over time, such as long-term adoption of ideas or technologies, might not be fully captured in a static analysis

For instance, in social media, influencers' roles evolve as user activity changes. Static metrics, while useful for a snapshot, may not capture long-term trends in influence propagation

Strong and Weak Ties in Networks

Critical Links in Information Flow

Identifying critical links in network helps in improving network and analyze the information flow in a network

- Bridging Communities: Links with high importance often connect different communities or clusters in a network. Removing such links can lead to the fragmentation of the network
- Information Flow: In communication networks, high betweenness links control the flow of information between different regions of the network. They are often bottlenecks or points of vulnerability
- In transportation networks, bridges or roads with high link betweenness are critical for connecting different regions. Disrupting these links can significantly affect the overall connectivity of the system

Link Betweenness

Link Betweenness: A measure of the importance of an edge based on how often it lies on the shortest paths between pairs of nodes

The betweenness centrality of a link e is the sum of the fraction of all-pairs shortest paths that pass through e

$$C_B(e) = \sum_{s \neq t} \frac{\sigma_{st}(e)}{\sigma_{st}}$$

Where σ_{st} is the total number of shortest paths from node s to node t, and $\sigma_{st}(e)$ is the number of those paths that pass through link e

Critical Links: High betweenness edges are critical for maintaining the flow of information in the network and bridging different communities

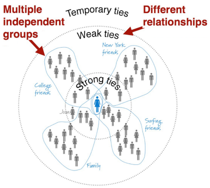
Bridges: An edge (i, j) is a local bridge if i and j have no common neighbors

Weak vs. Strong Ties in Networks

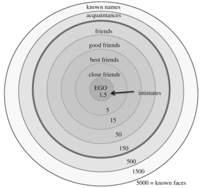
In many networks all edges are not the same

- Strong Ties: Close-knit relationships with frequent communication, emotional intensity, and dense connections (e.g., family, close friends)
- Weak Ties: More distant relationships with infrequent communication, lower emotional intensity, and greater potential to connect different social groups
- Both types of ties are essential for the overall functioning of social networks: strong ties ensure local cohesion, and weak ties extend the reach of information. In biological networks, tie strength can be based on biochemical interaction, in computer networks, it can be based on link bandwidth

Structure of human egocentric social networks



source: towardsdatascience.com



https://royalsocietypublishing.org/doi/10.1098/rspa.2020.0446

- Number of people in each circle increases, but contact frequency contact and closeness declines
- The outermost layer (5000) was identified by face recognition experiment (Num of faces that can be recognized as known by sight)

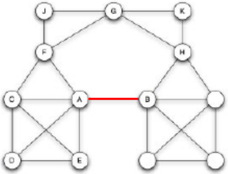
Strength of Weak Ties

Granovetter's Experiment: "Most job seekers (study subjects) found jobs through an acquaintance (weak tie) rather than a close friend (strong tie)"

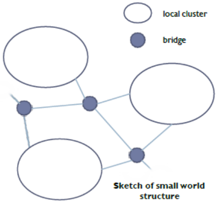
- Strong Ties for Cohesion: Information at end-points of a strong tie is nearly identical ▷ frequent synchronization
- Weak Ties for Information Diffusion: weak tie could help communication of novel information ▷ rare synchronization
- Role in Career Networking: Acquaintance can more likely inform of "new" job opportunities

Some edges that act as bridges between network segments, they are important for communication and explain the small-world phenomena in many networks

An edge (i, j) is a local bridge if i and j have no friends in common



source: Frank Dignum @ Umea University



Bridges and Weak Ties: Small-World and Information Spread

- Bridges: Edges that connect different network segments and play a significant role in maintaining network connectivity
- Small-World Networks: Bridges help explain the small-world phenomenon, where individuals can be connected in a small number of steps despite the network being sparse
- Granovetter's Theory: Weak ties as bridges are central to understanding the small-world property of social networks, where individuals can quickly access information from different communities

- Weak Ties in Information Spread: Granovetter's theory highlights how weak ties help spread information across a network by connecting individuals from different communities
- Access to Novel Information: Weak ties provide access to diverse sources of information, making individuals with many weak ties more influential in spreading new ideas
- Rapid Spread of Trends: Weak ties play a central role in enabling the rapid spread of trends, innovations, and viral content across a large network

Dynamics of Influence and Real-World Applications

Dynamic Influence Models

In most Real-World scenarios graphs are not static, the nodes (including any attributes associated with them), links and overall topology is constantly changing and evolving with time - Models need to cater to the evolving networks, particularly changes in node attributes

Influence diffusion models: These models describe how ideas, behaviors, or products spread through a network. There has been extensive research in this area and researchers have come up with several different models

Broad groups of these models based assumptions about how influence operates:

Epidemic-Based Models

- SIS Model (Susceptible-Infected-Susceptible)
- SIR Model (Susceptible-Infected-Recovered)

Activation-Based Models (Discrete Time)

- Independent Cascade (IC) Model
- Linear Threshold (LT) Model
- Threshold-Cascade (Granovetter's) Model

Opinion Dynamics Models

- DeGroot Model
- Voter Model
- Game-Theoretic Models

Continuous-Time Diffusion Models

- Hawkes Processes

Viral Marketing vs. Traditional Marketing and Recommendation

- Hotmail gained 18M users in 12 months, spending $50K on traditional ads
- Gmail rapidly gained users although referrals were the only way to sign up
- Google AdSense helps sellers reach buyers with targeted advertising
- Customers are becoming less susceptible to mass marketing
- Mass marketing is impractical for the huge variety of products online
- Over 50% of people do research online before purchasing electronics
- Personalized recommendations are based on prior purchase patterns and ratings

In social networks, users are more influenced by "friends" than by strangers

[Burke 2003]: 68% of consumers consult friends and family before purchasing home electronics

Viral Marketing (not Viralization/Viral Spreading) successfully utilizes social networks influence for adoption of some products/services

Applications of Dynamic Models in Social Media Analysis

Dynamic models are widely used to analyze and predict trends in social media, where network structures and interactions change rapidly:

- Influence Prediction: Dynamic models help predict which users are likely to become influencers as their connections grow over time
- Viral Content: They can effectively capture how content goes viral by tracking how influence cascades through the network in real-time
- Social Movements: Movements like the #MeToo or Black Lives Matter gain momentum dynamically as more users join the conversation and share content over time
- Brand Promotion: Companies can use appropriate models to track how advertising campaigns spread through social media, helping to optimize marketing strategies

Dynamic models provide more accurate and flexible tools for understanding and harnessing influence in fast-changing environments like social media.

Social Media Campaigns and Viral Marketing

Social media campaigns leverage the Diffusion of Innovations theory to rapidly spread content, products, or ideas across a wide audience. In viral marketing, the goal is to initiate a cascade of influence that reaches the early majority and beyond

Key factors in viral marketing:

- Influential Nodes: Identifying key individuals (early adopters, influencers) who can trigger widespread adoption
- Content Creation: Developing content that is engaging, shareable, and relevant to the target audience
- Timing: Launching campaigns at the right moment, when the audience is most likely to engage

Viral marketing campaigns exploit social networks to create exponential growth in awareness and adoption of products or ideas.

Epidemic-Based Models

Spread of Influence as an Epidemic

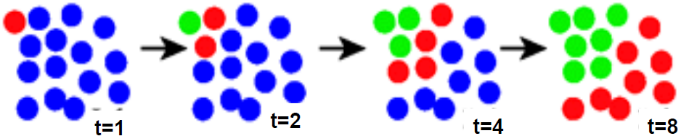
Two paradigms of Epidemiology:

- Microscopic: Researchers try to disassemble and neutralize new viruses ⇒ quest for vaccines, treatment, and cure
- Macroscopic: Statistical analysis and modeling of epidemiological data in order to find information and policies aimed at lowering epidemic outbreaks ⇒ macroscopic prophylaxis, containment and vaccination campaigns

Standard Epidemic Modeling: Groups of Subjects

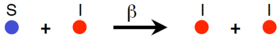
source: Alain Barrat @ Turin, Itlay

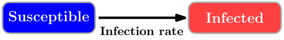
Neglecting differences in: age, gender, health, social class/status, susceptibility to disease, latency, severity of disease, ...



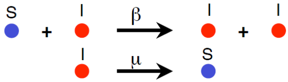
Standard Epidemic Modeling

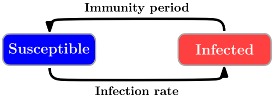
Epidemic models for the spread of “virus”, rumors, innovations, or technologies



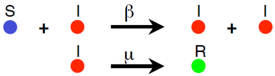


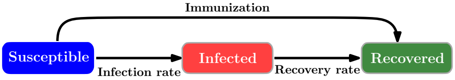
In SI, once a node is infected, it stays infected permanently, and the spread continues until all nodes in the network are infected





SIS is suitable for modeling recurrent influence (no permanent adoption)





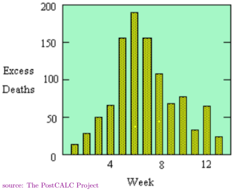
In SIR Infected individuals “can recover/are removed” and no longer participate in the spread, useful for modeling behaviors or trends with a limited lifespan

Hong Kong Flu Case Study: Modeling with SIR

In 1968-1969, the Hong Kong flu caused a significant epidemic

- Approximately 100,000 excess pneumonia and influenza deaths in NY City
- Data on weekly excess deaths provides insight into the epidemic dynamics
- Assume number of excess deaths in a week was proportional to the number of new cases of flu in say three weeks earlier
- Recovered group include dead, $\because$ they can't contract the disease
- Modeling outbreaks helps predict disease spread and control strategies

Week	1	2	3	4	5	6	7	8	9	10	11	12	13
Flu-Related Deaths	14	28	50	66	156	190	156	108	68	77	33	65	24



source: The PostCALC Project

Material on SIR Modelling is adapted from The PostCALC Project @ Duke University

Basic SIR Model Setup

The SIR model divides the population into three compartments:

- Susceptible (S): not yet infected,
- Infected (I): currently infected,
- Recovered (R): recovered or died (immune).

$$S \xrightarrow{\beta I} I \xrightarrow{\mu} R$$

- $\beta =$ transmission rate
- $\mu =$ recovery rate

$S(t), I(t), R(t) \quad \Rightarrow \quad$ number of individuals in each compartment at time t

Population assumed constant: $N = S(t) + I(t) + R(t)$.

$s(t) = \frac{S(t)}{N}, \quad i(t) = \frac{I(t)}{N}, \quad r(t) = \frac{R(t)}{N}$

Assumption: Homogeneous mixing; every infected individual has b effective contacts per day, not all with S, a fraction $s(t)$ with people in S

New infections $\propto b \times s(t) \times I(t)$ $\quad \triangleright$ each infected individual generates $bs(t)$ new infected individuals per day

Recovery rate: $\quad k =$ fraction of infected recovering per day

No one added to group S, since population is constant (no births/migration)

Differential Equations Formulation

Susceptible population changes: $\frac{ds}{dt} = -bs(t)i(t)$

Recovered population changes: $\frac{dr}{dt} = ki(t)$

Infected population changes:

$$\frac{ds}{dt} + \frac{di}{dt} + \frac{dr}{dt} = 0 \implies \frac{di}{dt} = bs(t)i(t) - ki(t)$$

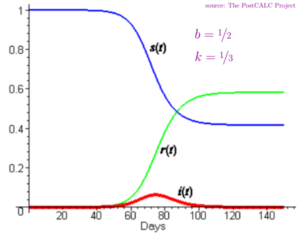
Initial conditions for the ODE system:

$$S(0) = 7.9M, \quad I(0) = 10, \quad R(0) = 0 \quad s(0) \approx 1, \quad i(0) = 1.27 \times 10^{-6}, \quad r(0) = 0$$

We need to estimate the value of k and b

Complete ODEs model:

$$^{ds}/_{dt} = -b\,s(t)\,i(t), \quad s(0) = 1,$$
$$^{di}/_{dt} = b\,s(t)\,i(t) - k\,i(t), \quad i(0) = 1.27 \times 10^{-6},$$
$$^{dr}/_{dt} = k\,i(t), \quad r(0) = 0$$



Numerical Solutions: Euler's Method

Euler's method approximates the system discretely:

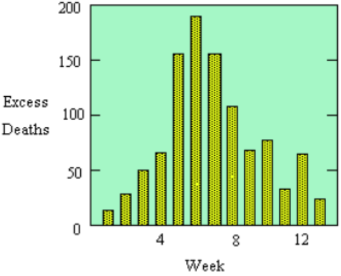
$$
\begin{aligned}
s_n &= s_{n-1} - b\, s_{n-1} i_{n-1} \Delta t, \\
i_n &= i_{n-1} + (b\, s_{n-1} i_{n-1} - k\, i_{n-1})\, \Delta t, \\
r_n &= r_{n-1} + k\, i_{n-1} \Delta t
\end{aligned}
$$

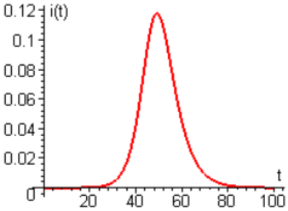
To use these formulae, we need explicity values for b, k, $s(0)$, $i(0)$, $r(0)$, and t

- Δt is the step size, Smaller Δt increases accuracy
- Simple and effective for simulating epidemic curves
- Euler's method enables numerical simulation when analytic solutions are unavailable
- Numerical results can be compared to real data
- Find a step size for which the Euler solutions appear to closely track true solutions of the system

Parameter Sensitivity and Data Fitting

- Infectious period estimated as 3 days $\Rightarrow k = 1/3$
- b is estimated by fitting the model to death data
- Varying b affects the infection curve $i(t)$:
  - Larger b increases peak infected fraction and speed of spread
  - Smaller b slows epidemic and lowers peak
- Varying k alters duration of infection and peak height
- Model with $b = 0.6$, $k = 1/3$ fits NYC flu death data reasonably well





source: The PostCALC Project

The Contact Number

c = average number of close contacts per infected individual during infectious period, meansures disease contagiousness relative to population mixing

$$c = \frac{b}{k}$$

By chain rule we have $\frac{di}{ds} = \frac{di/dt}{ds/dt}$

Recall $di/dt = b\,s(t)\,i(t) - k\,i(t)$ and $ds/dt = -b\,s(t)\,i(t)$ we get

$$\frac{di}{ds} = -1 + \frac{1}{cs}$$

This determines (except for dependence on an initial condition) the infected fraction i as a function of the susceptible fraction s

The solution yields: $i(s) = 1 - s + \frac{1}{c}\ln\left(\frac{s}{s_0}\right)$, where s_0 is initial susceptible fraction

Using $s(\infty)$ (final susceptible fraction), c can be estimated from epidemic data

Herd Immunity and Vaccination

Herd immunity occurs when enough individuals are immune (through previous infections or vaccination) to prevent epidemic spread

Vaccination reduces the susceptible population, contributing to herd immunity

▷ Vaccination introduces direct transition: $S \to R$ bypassing infection

Recall the Contact Number $c = {}^{b}/_{k}$

If $s_0 < {}^{1}/_{c}$, then ${}^{di}/_{dt}$ remains negative, preventing the disease from spreading, and an epidemic cannot occur ▷ Threshold condition for herd immunity

$$\frac{di}{dt} = bs(t)i(t) - ki(t) = i(t)\big(bs(t) - k\big)$$

The factor $i(t)$ is always non-negative, for ${}^{di}/_{dt} < 0$, we require $s(t) < {}^{k}/_{b} = {}^{1}/_{c}$

Thus, reducing $s(t)$ through vaccination prevents the epidemic.

To achieve herd immunity through vaccination, we need $v > 1 - {}^{1}/_{c}$, where v is the fraction of the population that must be vaccinated, for example

- If $c = 12.8$, the fraction needed is $v > 1 - {}^{1}/_{12.8} \approx 0.92$ or 92%
- If the vaccine is only 95% effective, a higher fraction is required

SIR Model

The SIR model is particularly effective in measuring the life cycle of viral content on social media:

- Viral Spread: Content spreads rapidly when it is first shared, infecting a large number of users (adopters)
- Peak Influence: After reaching a peak, the spread begins to slow down as more individuals become "recovered" and no longer share or engage with the content
- Content Saturation: Eventually, the spread stops as the entire population is either recovered or unaffected

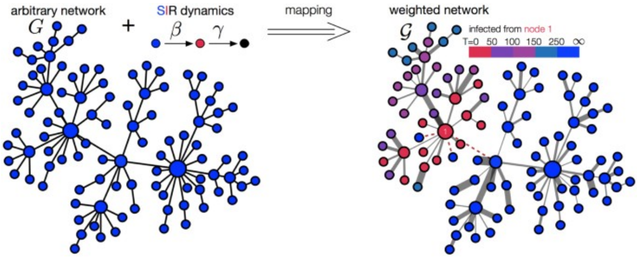


Figure: SIR model example, starting from a single infected node at $t = 0$

Basic SI Model Setup

The SIR model divides the population into two compartments:

- Susceptible (S): not yet infected,
- Infected (I): currently infected,

$$S \xrightarrow{\beta I} I$$

- $\beta =$ transmission rate

Example: The SI model can be used to predict the spread of a viral video; I : those who have seen and shared the video, and S : those who have not seen it yet

$S(t), I(t) \quad \Rightarrow \quad$ number of individuals in each compartment at time t

Constant population: $N = S(t) + I(t) \qquad s(t) = \frac{S(t)}{N}, \quad i(t) = \frac{I(t)}{N}$

Assumption: Homogeneous mixing; every infected individual has b effective contacts per day, not all with S, a fraction $s(t)$ with people in S

$$\begin{aligned}\Pr(S \to I) &= 1 - \Pr(\text{not infected by any infectious}) \\ &= 1 - (1 - \beta dt)^{ki(t)} \approx \beta ki(t)dt \quad (\beta dt \ll 1)\end{aligned}$$

$$\frac{di}{dt} = s(t) \times \Pr(S \to I) = \beta ki(t)s(t) = \beta ki(t)(1 - i(t))$$

Recurrent Influence and the SIS Model

The SIR model divides the population into two compartments:

- Susceptible (S): not yet infected,
- Infected (I): currently infected,
- After some time, infected individuals can become susceptible again

$$S \xrightarrow{\beta} I \xrightarrow{\mu} S$$

- β = transmission rate
- μ = recovery rate

Used to predict virus with no permanent immunization, it can also model rumor, habits, modeling recurrent influence (no permanent adoption), product adoption (subscribe, unsubscribe, and potentially re-subscribe later)

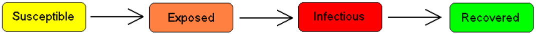
$S(t), I(t) \quad \Rightarrow \quad$ number of individuals in each compartment at time t

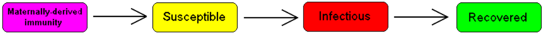
Constant population: $N = S(t) + I(t) \qquad s(t) = \frac{S(t)}{N}, \quad i(t) = \frac{I(t)}{N}$

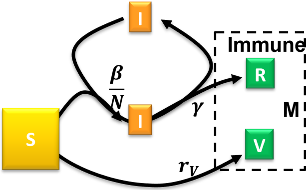
- $\frac{di(t)}{dt} = \beta s(t)i(t) - \mu i(t) = \beta i(t)(1 - i(t)) - \mu i(t)$
- $\frac{ds(t)}{dt} = -\beta s(t)i(t) + \mu i(t) = -\beta i(t)(1 - i(t)) + \mu i(t)$

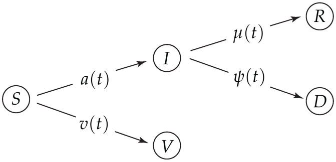
Other compartmental models

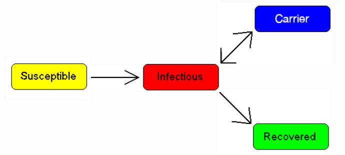
- SIRD
- SIRV
- SIRVD
- SIRVB
- SIRS
- MSIR
- Carrier
- SEIR
- SEIS

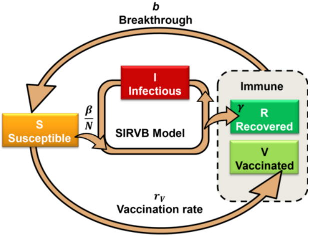












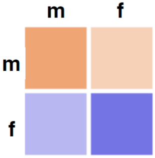
source: Wikipedia

Incorporating Population Structure and Non-Homogeneous Mixing

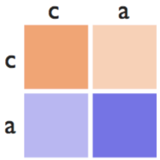
Different classes of individuals (age, gender, etc.) lead to potentially different:

- Transmissibility
- Contact rates

Adapted from Alain Barrat @ Turin, Itlay

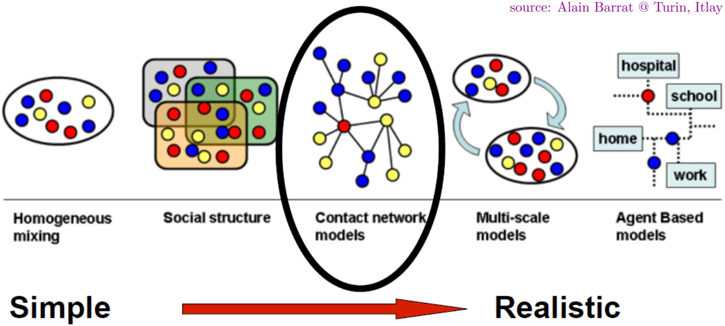


HIV transmission differs depending on gender



Flu contact rates are different for adults and children

source: Alain Barrat @ Turin, Itlay



Activation-Based Models

Independent Cascade (IC) Model

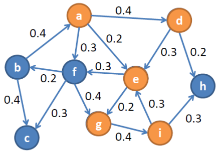
The IC model is a stochastic process where nodes in a network can activate their neighbors with a given probability. This model is important because:

- It simulates viral marketing campaigns and information diffusion
- It is used to predict the spread of diseases or rumors
- The IC model can predict how a new product may spread on Twitter or Facebook
- It helps design strategies to maximize influence in social networks

Independent Cascade (IC) Model

IC: A discrete-time probabilistic model of diffusion in an edge weighted network G

- $p(u,v) = p_{uv}$ is probability that node u will influence v
- Initially some nodes are active and activations spread in G
- Each edge (u,v) has probability (weight) p_{uv}
- Each active node u has a **single chance** to activate each inactive neighbor v with probability p_{uv}
- (Independent) activation attempts happen in next time step



Let $G = ((V,E),p)$ be a digraph. $p : E \mapsto [0,1]$

- $A_0 \subseteq V$: initial active set $\triangleright$ (seed nodes)
- At time t, each $u \in A_t \setminus A_{t-1}$ tries to activate each $v \in N(u)$ with prob. p_{uv}
- If successful, v is added to A_{t+1}

$$\mathbb{P}[v \text{ activated at } t+1] = 1 - \prod_{u \in N(v) \cap A_t} (1 - p_{uv})$$

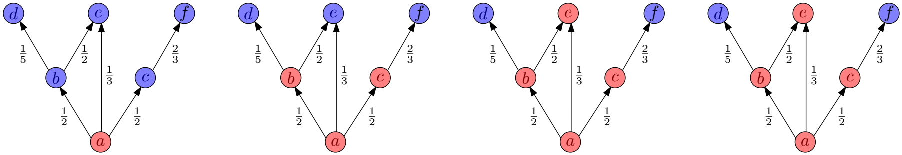
Independent Cascade (IC) Model

IC: A discrete-time probabilistic model of diffusion in an edge weighted network G

Let $G = ((V, E), p)$ be a digraph. $p : E \mapsto [0, 1]$

- $A_0 \subseteq V$: initial active set ▷ (seed nodes)
- At time t, each $u \in A_t \setminus A_{t-1}$ tries to activate each $v \in N(u)$ with prob. p_{uv}
- If successful, v is added to A_{t+1}

$$\mathbb{P}[v \text{ activated at } t+1] = 1 - \prod_{u \in N(v) \cap A_t} (1 - p_{uv})$$



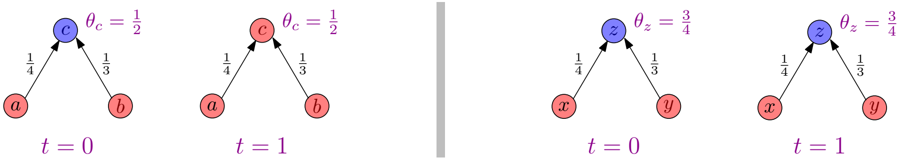
$t = 0$ $t = 1$ $t = 2$ $t = 3$

Linear Threshold Model

Given an edge-weighted (undirected or directed) network

- Edges weight, $w(u,v) = w_{uv} \in [0,1]$ is the relative influence of node u on node v (e.g., quantitative version of weak and strong ties)
- Each node v independently selects a threshold $\theta_v \sim$ Uniform$[0,1]$ to model uncertainty about how easy it is to influence that individual

  ▷ Prior knowledge of the types of nodes may better inform the distribution from which θ_v is sampled
- A node v becomes active at time t if: $\sum_{u \in A_{t-1}} w_{uv} \geq \theta_v$
- Once activated, nodes remain active permanently

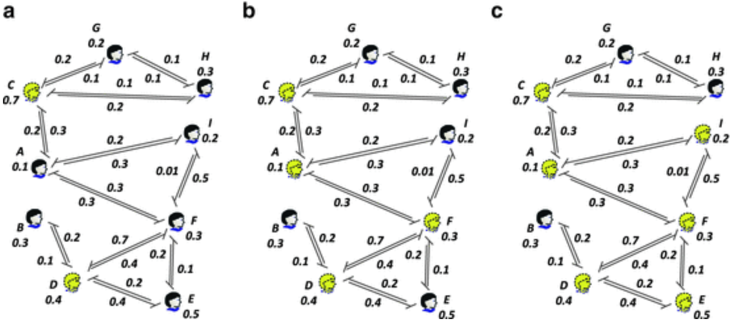


Independent Cascade vs. Linear Threshold Model

The LT model has similarities with the IC model with the following exceptions:

- Nodes also have an intrinsic attribute for their threshold θ, for example $\theta_C = 0.7$ and $\theta_I = 0.2$
- Unlike the IC model, activated nodes can continue to try to influence their neighboring nodes at each time step

If at any time step the sum of the influence of all activated neighbours for a node v exceeds its threshold then the node will get activated: $\sum_{u \in A_{t-1} \cap N_v} w_{uv} \geq \theta_v$



Granovetter's Threshold Model

Activation condition for Granovetters model

A node v becomes active at time t if the *fraction* of its neighbors that are active at time $t-1$ exceeds its threshold θ_v:

$$\frac{||N_v \cap A_{t-1}|}{|N_v|} \geq \theta_v$$

- N_v: Set of neighbors of node v
- A_{t-1}: Set of active nodes at time $t-1$
- $\theta_v \in [0,1]$: Fractional threshold for activation

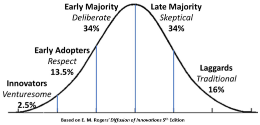
- Unlike the Linear Threshold Model, where weighted influence matters, Granovetter's model uses purely the fraction of active neighbors. In that sense it is a bit like a simplified version of the Linear Threshold model
- It's useful for modeling social contagion based on peer proportions (e.g., protests, adoption of norms)

Thresholds and Social Adoption Types

In practice, based on sociological theory most individuals can be categorised into different groups depending on their affinity to get influenced or adopt innovations

Especially in threshold-based diffusion models (e.g., LT, Granovetter), a node's threshold $\theta_v \in [0, 1]$ determines its resistance to adopting a new behavior

- Innovators: $\theta_v \approx 0$ — adopt early, even with little or no peer support
- Early Adopters: $\theta_v \in (0, 0.25]$ — adopt quickly with minimal social proof
- Early Majority: $\theta_v \in (0.25, 0.5]$ — need moderate peer adoption
- Late Majority: $\theta_v \in (0.5, 0.75]$ — adopt only after most others have
- Laggards: $\theta_v \in (0.75, 1]$ — very resistant to change; require overwhelming influence



Such categorization can also apply to how much influence a node exerts. In LT model

- Outgoing edge weights w_{uv} represent the influence of node u on neighbor v
- High-profile individuals (celebrities, experts) have higher w_{uv} to many nodes — they are strong spreaders
- Ordinary individuals contribute less influence per edge

Opinion Dynamics in Social Networks

DeGroot Model of Opinion Dynamics

Model Overview

The DeGroot model is *discrete-time*, *synchronous* model of opinion diffusion, where agents update their beliefs as weighted averages of their neighbors' opinions

- Each node v has a scalar opinion $x_v(t) \in [0, 1]$ at time t
- The opinion at the next time step is:

$$x_v(t+1) = \sum_{u \in N_v} w_{uv} \cdot x_u(t)$$

- The weights w_{uv} form a row-stochastic matrix (i.e., $\sum_u w_{uv} = 1$)

Goal: Model how consensus or polarization emerges over time

- The weights w_{uv} capture how much trust or importance node v places on node u's opinion
- High-influence individuals (leaders, experts) will have large outgoing weights across the network
- Nodes with stubborn or fixed opinions can be modeled by assigning them a self-loop with high weight

Sociological Insights from the DeGroot Model

- Consensus: If all nodes iteratively average, the network may converge to a shared opinion
- Polarization: With certain network structures or stubborn nodes, multiple persistent opinions can coexist
- Echo Chambers: Can arise when subgroups mostly weight each other

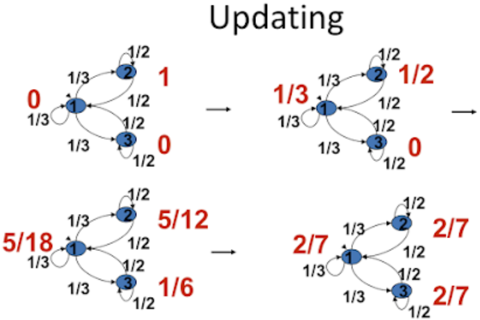


Figure: How opinions are updated at each time step in the DeGroot model

Voter Model of Opinion Dynamics

Model Overview

The Voter model is a *discrete-time*, *randomized* model of opinion dynamics, where agents adopt the opinion of a randomly selected neighbor at each time step

- Each node v has a binary opinion $x_v(t) \in \{0, 1\}$ at time t (Could be a larger set)
- At each time step, a node v randomly selects a neighbor $u \in N_v$ and adopts their opinion:

$$x_v(t+1) = x_u(t)$$

- This process is repeated asynchronously or synchronously across the network

Goal: Model how consensus or majority opinion forms in a population over time

- In the simplest case, nodes are equally likely to interact with any of their neighbors
- The network's structure influences the spread of opinions (e.g., small-world, random graphs)
- Fixed opinions or external influences can be introduced to simulate stubborn agents

Sociological Insights from the Voter Model

- Consensus: In many scenarios, the network will eventually reach a single, unanimous opinion. If the network has a large number of distinct, fixed opinions or stubborn agents, polarization may persist
- Majority Influence: A majority opinion can dominate the network, depending on the initial distribution of opinions
- Coexistence of Opinions: Small initial minority groups may persist, especially in sparsely connected networks or when external biases exist

The voter model is significantly dependent on the topology and structure of the graph

- Cliques in the graph may result in a very strong localized opinion since each of the nodes in the clique are reinforcing each other opinions
- Cycles, especially those with an even number of nodes, can lead to a situation where nodes alternate in opinion, preventing the system from reaching consensus quickly. This results in a kind of "opinion oscillation"
- A bridge is an edge in the graph whose removal would disconnect the network into two or more components. If a bridge exists between two large components, it can prevent the spread of opinions between them. The two components may converge to different opinions

Other components may have similar disproportionate effects on the spread of opinions

Game-Theoretic Models of Opinion Dynamics

Opinion dynamics can be framed as strategic interactions, where agents maximize payoffs influenced by their own preferences and the opinions of their neighbors.

These models capture rational or boundedly rational behavior

- Each agent i selects an action or opinion $s_i \in S$ to maximize a utility function $u_i(s_i, s_{-i})$
- Equilibria (e.g., Nash or logit equilibrium) model steady-state opinions.
- Game-theoretic models can handle both discrete (binary/categorical) and continuous opinion spaces
- These models go beyond pure diffusion by including *strategic reasoning*, *inertia*, and *heterogeneity*

Setup for Game-Theoretic Models

General Formulation of Game-Theoretic Opinion Dynamics

Let $G = (V, E)$ be a graph where each node $i \in V$ holds an opinion or strategy $s_i \in \mathcal{S}$, and has a utility function $u_i : \mathcal{S}^n \to \mathbb{R}$ that depends on both its own strategy and those of its neighbors

- Agents update their strategy via a noisy best-response:

$$\mathbb{P}(s_i(t+1) = a) \propto \exp\left(\frac{u_i(a, s_{-i}(t))}{\tau}\right)$$

- Alternatively, agents minimize local cost:

$$s_i(t+1) = \arg\min_{a \in \mathcal{S}} \left[c_i(a, s_{-i}(t)) + \eta_i(a)\right]$$

The functions u_i or c_i can encode coordination, conformity, stubbornness, or polarization depending on network structure and agent biases. The parameter τ or penalty η_i controls rationality or noise

Potential and Graphical Games in Opinion Dynamics

Local Interactions and Stability

In potential and graphical games, each agent's payoff depends only on a subset of neighbors, and dynamics can often be tracked via a global potential function.

- In a potential game, there exists Φ such that:

$$u_i(s_i', s_{-i}) - u_i(s_i, s_{-i}) = \Phi(s_i', s_{-i}) - \Phi(s_i, s_{-i})$$

- Best-response dynamics increase Φ and often converge
- In graphical games, utility of i depends only on $N(i)$, its neighborhood
- Models like the Friedkin–Johnsen or threshold models can be interpreted as such games

Log-Linear Learning and Bounded Rationality

Behavioral Realism via Noisy Best-Response

Log-linear learning captures noisy decision-making where agents select actions with probability proportional to their payoffs

- Action probabilities follow a Gibbs distribution:

$$\mathbb{P}(s_i = a) \propto \exp\left(\frac{u_i(a, s_{-i})}{\tau}\right)$$

- As temperature $\tau \to 0$, dynamics converge to pure best response
- System tends to stochastically stable equilibria of the potential game
- Captures bounded rationality, exploration, and noise in human behavior

Continuous Opinion Games & Cost Minimization

Opinion as Continuous Strategy

Continuous opinion models define agents' costs as functions of disagreement with neighbors and deviation from intrinsic bias

- Example: Friedkin–Johnsen model minimizes:

$$C_i(x) = \sum_{j \in N(i)} w_{ij}(x_i - x_j)^2 + \lambda_i(x_i - s_i)^2$$

- Unique equilibrium exists under convexity
- Can be framed as potential games with convergence under best-response
- Also interpretable as constrained optimization under bounded confidence (e.g., Hegselmann–Krause)

Dynamics and Emergent Behavior

Game-theoretic models reveal how equilibrium behavior reflects social structure, individual incentives, and learning processes.

- Consensus: arises under aligned incentives or strong coordination
- Polarization: emerges when agents have diverging biases or localized trust
- Stubbornness: modeled via self-preference terms or fixed players
- Stability: convergence guaranteed in potential games, often under asynchronous or noisy updates

Often more fine-grained analysis and detail then models like DeGroot

Game theoretic models more effectively capture the intricacies involved in strategic interaction, but at the cost of significantly more complex models

When modelling simpler scenarios these may overcomplicate analysis

Continuous-Time Diffusion Models

Continuous-Time Diffusion Models

Motivation

Unlike discrete-time models (e.g., DeGroot or Voter), continuous-time models capture *when* events happen, allowing for fine-grained modeling of diffusion in real-world settings (e.g., retweets, citations)

- Nodes in a network influence each other via random events over time
- The intensity of events depends on past history — more recent events often lead to higher likelihood of further diffusion
- These models are especially suited for settings where timing matters: social media, news spread, epidemics

Example Models: Poisson processes, Self-exciting point processes (Hawkes), Survival analysis models.

Hawkes Process for Diffusion

Mathematical Formulation

Let $\lambda_i(t)$ be the intensity (rate) at which node i generates events (e.g., adopts an opinion, shares information). Then:

$$\lambda_i(t) = \mu_i + \sum_{j \in N_i} \sum_{t_k^j < t} \alpha_{ij} \cdot g(t - t_k^j)$$

- μ_i is the base rate of spontaneous activation
- $\alpha_{ij} \geq 0$ is the influence of node j on node i
- $g(\cdot)$ is a kernel function (e.g., exponential decay: $g(t) = \beta e^{-\beta t}$)

- Hawkes processes are *self-exciting*: one event increases the likelihood of future ones
- Can model cascades, virality, and temporal clustering of opinions or behaviors
- Extensions: multivariate, marked, nonlinear Hawkes for richer dynamics

Influence Maximization

Influence Maximization Algorithms

Influence Maximization refers to strategies used to identify the most influential nodes in a network, in order to maximize the spread of influence within a given budget or time constraint.

Problem Definition

Given a diffusion model (e.g., LT, IC), a graph $G = (V, E)$, and a budget k, find a seed set $S \subseteq V$, $|S| = k$, that maximizes the expected spread of influence $\sigma(S)$.

$$S^* = \arg \max_{S \subseteq V,\ |S| = k} \sigma(S)$$

- **Submodularity:** For many models (e.g., IC, LT), $\sigma(\cdot)$ is monotonic and submodular

  $\therefore$ A greedy approximation gives a $(1 - 1/e)$-approximation.

- **Greedy Approach:** Iteratively select the node v with the largest marginal gain:

$$v = \arg \max_{u \in V \setminus S} \sigma(S \cup \{u\}) - \sigma(S)$$

Strategies to Identify the Most Influential Nodes

Improvements: There are several algorithms for selecting the initial seed set more effectively. These may leverage information about the social network or graph structure such as:

- Community Structure: Selecting seeds from different communities to maximize inter-group spread.
- Node Centrality: Prioritizing nodes with high degree, betweenness, or eigenvector centrality.
- Bridge Nodes: Targeting nodes that act as bridges between densely connected clusters.
- Influence Estimation Models: Using diffusion simulations (e.g., Monte Carlo) or RR set sampling (e.g., IMM) to estimate marginal influence.
- Historical Data: Learning influence probabilities from past diffusion traces to inform seed selection.

Influence Maximization in Activation-Based (Discrete Time) Models

Problem:

Given a diffusion model (IC or LT), find a seed set $S \subseteq V$ of size k that maximizes expected influence spread $\sigma(S)$.

Approaches:

- Greedy Algorithm: Classic method using Monte Carlo simulations. Guarantees $(1 - 1/e - \varepsilon)$-approximation.
- CELF/ CELF++: Speed up greedy by exploiting submodularity via lazy evaluations.
- IMM / TIM / OPIM: Use Reverse Reachable (RR) sets to reduce to a maximum coverage problem.

Assumptions: Static graph structure; influence probabilities or weights are known.

These algorithms provide theoretical guarantees but vary in scalability and model-specific adaptations.

Influence Maximization in Epidemic Models

Model:

Epidemic dynamics such as SIR/SIS define probabilistic transitions between states (Susceptible, Infected, Recovered).

Algorithmic Strategies:

- Degree Heuristics: High-degree nodes are more likely to spread infection.
- Simulated Annealing / Genetic Algorithms: Optimize seed set under stochastic simulations of SIR
- NetShield (Kempe et al.): Minimize vulnerability or maximize control in epidemic outbreaks

Challenges: No submodularity in SIR, leading to weaker approximation guarantees

Epidemic-aware influence maximization emphasizes risk reduction and outbreak prediction

Game-Theoretic Influence Maximization

Setting:

Nodes are strategic agents choosing opinions or actions to maximize individual utility. Diffusion emerges from equilibrium behavior

Algorithmic Methods:

- Best-Response Dynamics: Simulate influence under equilibrium convergence
- Log-Linear Learning: Add noise to choices, derive probabilistic adoption influenced by potential functions
- Max-Potential Heuristics: Seed nodes to guide the system to a desirable equilibrium state

Remarks: Influence maximization is no longer purely additive — strategic externalities and rationality constraints dominate

Suitable for modeling rational agents in economic or political systems

Influence Maximization in Continuous-Time Models

Model:

Diffusion unfolds asynchronously over continuous time, e.g., via Hawkes Processes or Continuous-Time IC (CTIC)

Methods:

- NetRate: Learn edge transmission functions from data, then simulate diffusion for seed selection
- ConTinEst: Estimate influence spread under CTIC efficiently using stochastic shortest paths
- CMAB-based Methods: Use bandit algorithms to balance exploration and exploitation in unknown environments

Advantages: Models fine-grained temporal dynamics and heterogeneous edge delays

More realistic for viral marketing, information cascades, or rumor spreading in fast-moving environments

Influence in Opinion Dynamics Models

Setting:

Nodes update scalar-valued opinions over time via averaging (e.g., DeGroot) or best response to neighbors

Control Algorithms:

- Leader Selection: Choose stubborn agents (zealots) that anchor opinions to desired targets
- Anchoring Control: Optimize weights or placements of anchors to maximize final consensus
- Influence Shaping (Friedkin-Johnsen): Inject minimal control to steer the population opinion

Goal: Influence steady-state or convergence trajectory of opinion vectors $x(t) \in [0, 1]^n$

These models excel at capturing gradual influence, resistance, and convergence behaviors

Beyond Classical Seed Selection

Traditional influence maximization focuses on choosing a fixed set of k seed nodes. However, real-world settings often demand more flexible approaches:

- Adaptive Seeding: Seeds are selected sequentially across multiple rounds, observing partial spread before selecting the next
- Budgeted Influence: Nodes have varying costs. The goal is to maximize spread under a total cost constraint, not a fixed seed count
- Time-Sensitive Maximization: Maximize influence *by a deadline*, not eventual spread

Mathematically:

$$\max_{S \subseteq V,\ \text{cost}(S) \leq B} \mathbb{E}[f(S, T)]$$

where $f(S, T)$ is the spread function by time T, and B is the budget

Alternative Influence Objectives

Influence is not always about maximizing raw spread:

- Targeted Influence: Maximize influence within a specific subpopulation (e.g., swing voters, vulnerable communities)
- Spike Adoption:
  - Maximize the initial burst or "spike" of adoptions at a particular moment in time
  - Example: Launching a product with a viral marketing campaign that generates an immediate wave of interest
  - Strategy: Target a seed set that will quickly trigger a rapid cascade, maximizing early-stage engagement or adoption
- Long-Term Influence:
  - Aim to sustain influence or maintain adoption over an extended period, rather than just maximizing short-term spread
  - Example: In political campaigns, influencing long-term voter loyalty, or in public health, ensuring consistent healthy behaviors over time
  - Strategy: Choose influencers or adopters who can maintain and reinforce their influence over time, rather than triggering immediate short-term effects

Adversarial Influence and Strategic Interference

Influence maximization often assumes a benign planner, but real-world settings involve strategic **adversaries** aiming to counter or manipulate spread:

- Disruption of Consensus: Adversaries may aim to delay or prevent convergence in opinion dynamics (e.g., Voter, DeGroot models)
- Spread of Misinformation: Deploying seed nodes to maximize the reach of fake news or divisive content
- Blocking Influence Paths: Removing or corrupting nodes/edges to reduce the effectiveness of a seed set
- Targeted Deception: Influencing specific subgroups with tailored messages to skew public sentiment

Example Formulation:

$$\min_{A \subseteq V, |A| \leq k} \mathbb{E}[f(S \setminus A)]$$

where A is the set of disrupted nodes and S is the seed set
Game-theoretic models naturally capture such adversarial settings (Stackelberg games, zero-sum diffusion games)

Game-Theoretic Models of Adversarial Influence

Strategic interactions between a planner (defender) and an adversary can be modeled using:

- Stackelberg Games: The defender (leader) commits to a seed set; the attacker (follower) reacts by targeting nodes to disrupt
- Zero-Sum Games: Defender and attacker have opposing utility functions over influence spread
- Diffusion Games: Multiple agents (e.g., brands, parties) simultaneously seed opinions and compete for network influence

Stackelberg Formulation:

$$\max_{S:|S|\leq k} \min_{A:|A|\leq b} \sigma(S \setminus A)$$

where S is the seed set, A is the set of blocked or attacked nodes, and $\sigma(\cdot)$ denotes the expected spread.

Extensions:

- Robust seed selection under uncertainty or noise
- Time-aware defenses: spreading before or after attacks
- Detection games: identifying adversarial nodes via observation

Influence in Social Media Platforms

Social media platforms such as Twitter, Facebook, and YouTube deal with influence dynamics every day. The main goals are:

- Maximizing User Engagement: Ensuring content spreads effectively to engage users and keep them active
- Recommendation Systems: Algorithms aim to influence users' content consumption to increase platform time
- Controlling Misinformation: Balancing influence maximization with controlling the spread of harmful content
- Filter Bubbles: Avoiding the over-targeting of content based on users' past behavior, which leads to echo chambers

Some of the main concerns that arise are:

- Ethical concerns over manipulation
- Identifying harmful or malicious influence

Spread of Misinformation and Fake News

One of the most critical issues for social media platforms is the spread of misinformation:

- Fake News: The rapid spread of false or misleading information, often designed to manipulate public opinion
- Adversarial Influence: Malicious actors spreading divisive content to influence elections or public sentiment
- Viral Content: High engagement rates for controversial or emotional content, even when harmful

Challenges:

- Identifying credible vs. unreliable sources
- Managing viral misinformation in real-time
- Developing scalable fact-checking mechanisms

Impact:

- Threatens public trust and social cohesion
- Can lead to polarization and unrest

Adversarial Influence and Manipulation

Social networks face strategic interference from malicious users who manipulate influence for personal or political gains:

- Astroturfing: Creating fake grassroots movements to simulate genuine public support for a cause
- Bot Networks: Using automated accounts to spread messages or influence discussions on a mass scale
- Deepfakes and Fake Content: Manipulating videos or images to deceive users and sway opinions

Challenges:

- High sophistication of adversaries
- Difficulty in distinguishing between genuine user behavior and manipulation

Solutions:

- Bot detection algorithms
- Deep learning models for identifying synthetic content
- Behavioral analysis to identify coordinated influence campaigns

Controlling Echo Chambers and Filter Bubbles

Echo chambers and filter bubbles result from algorithmic recommendations that reinforce existing beliefs:

- Echo Chambers: Groups of users who are exposed only to information that aligns with their views
- Filter Bubbles: Algorithms tailor content to users' preferences, leading them to miss out on diverse perspectives
- Polarization: Repeated exposure to like-minded content can intensify political or ideological divisions

Challenges:

- Balancing personalized recommendations with diversity of viewpoints
- Reducing algorithmic bias in content curation
- Encouraging healthy discussions and reducing divisive content

Solutions:

- Introducing content diversity mechanisms
- Designing recommendation systems that expose users to opposing viewpoints

Regulating Social Media Influence

Governments and social networks themselves are exploring regulations to manage influence:

- Transparency Laws: Requiring platforms to disclose how content is promoted and how influence spreads
- Accountability for Ads: Ensuring that political ads are clearly labeled and that their sources are identified
- Anti-Manipulation Efforts: Legal and technical mechanisms to detect and prevent harmful manipulation

Challenges:

- Balancing regulation with freedom of expression
- Addressing cross-border issues with global platforms

Solutions:

- Collaborations between tech companies, governments, and fact-checking organizations
- Enhanced algorithms for automatic detection of harmful influence tactics

Leveraging Influence Maximization for Product Adoption

Social media platforms and companies increasingly leverage influence maximization techniques to drive product adoption:

- Influencer Selection: Identifying influential figures who have the highest impact on their followers. This can involve:
  - Using network centrality measures to find key nodes (influencers) within the social graph
  - Evaluating audience demographics and alignment with the product
- Targeted Campaigns: Utilizing data from influence models to optimize advertising spend:
  - Selecting influencers whose audiences are most likely to adopt the product
  - Running personalized ads based on past user behavior and social connections
- Viral Marketing: Focusing on the early adopters who can trigger widespread influence and further adoption

Example: A company launching a new fitness product may partner with fitness influencers who have strong connections within the health and wellness community

Selecting Social Media Influencers for Maximum Reach

Companies often leverage social network models to maximize product reach through influencers:

- Centrality-Based Selection:
  - Finding influencers with the highest degree centrality (i.e., those with the most connections)
  - Analyzing betweenness centrality to identify individuals who bridge different subgroups
- Cascade Simulations:
  - Running simulations to model how information (e.g., a product launch) spreads across the network
  - Selecting influencers whose impact spreads the fastest across a network (using diffusion models like IC or LT)
- Targeting Early Adopters:
  - Identifying early adopters who are more likely to adopt a product quickly and influence their social circle
  - Using k-core or community detection to find groups likely to engage with the product

Example: A fashion brand may select influencers based on centrality in a fashion-related social network, targeting individuals who will lead to the fastest product spread

Political Campaigns: Maximizing Influence for Voter Outreach

Political campaigners apply influence maximization techniques to reach voters effectively and efficiently:

- Targeted Messaging:
  - Analyzing the social network of voters to identify key influencers and target them with personalized messages
  - Running simulations to identify which individuals or groups are most likely to influence others
- Key Opinion Leaders:
  - Identifying local leaders, community activists, or online influencers who have a significant impact on voter sentiment
  - Using social media monitoring to find high-influence individuals within specific voter demographics
- Message Amplification:
  - Using targeted ads and content to amplify messages among specific voter groups (using influence maximization strategies like greedy algorithms)
  - Leveraging opinion dynamics models to understand how messages spread and reinforce voter opinions

Example: During elections, campaigns target key swing voters and influencers within communities to sway voter sentiment and increase turnout

Competitive Influence Maximization: Political Campaigns

In a competitive influence maximization setting, agents (e.g., political campaigns) aim to sway the opinions of voters while countering the influence of their opponents

- Two Competing Agents: Each political party seeks to maximize their own voter base while minimizing the opponent's influence.
- Objective:
  - Maximize the spread of influence (voter support) within a targeted group of voters
  - Simultaneously minimize the opponent's reach by preventing adoption in certain groups
- Game-Theoretic Approach: Use of strategic planning to anticipate the opponent's actions and counter them effectively
- Strategy:
  - Identify influential nodes (voters or regions) and target them for early adoption
  - Disrupt the opponent's potential spread by targeting key adversary nodes or preventing adoption in regions of strategic importance

This setup requires dynamic game-theoretic strategies, where each agent adjusts their strategy in response to the opponent's actions

Competitive Influence Maximization: Competing Products

Similar to political campaigns, competitive influence maximization is essential for marketing competing products, where each brand aims to increase its market share while reducing the influence of its competitors

- Two Competing Products: Each company wants to maximize the adoption of its product while preventing the other's product from gaining traction
- Objective:
  - Maximize the adoption of the product within a target demographic
  - Minimize the spread of the competing product by blocking its reach in key market segments
- Strategic Use of Resources:
  - Efficient allocation of limited marketing resources (ad spend, influencer partnerships, etc.) to both spread influence and diminish the opponent's impact
- Game-Theoretic Considerations: Each company must anticipate the actions of its competitor, such as targeting the same demographic or engaging similar influencers

Companies must not only maximize their influence, but also react in real-time to their competitors' actions to maintain their edge

Challenges in Leveraging Influence Maximization

While influence maximization techniques provide powerful tools for adoption, there are significant challenges:

- Ethical Concerns:
  - Ensuring that influence techniques do not manipulate users in unethical ways
  - Addressing concerns over privacy and transparency in influence algorithms
- Detection of Manipulation:
  - Identifying and combating malicious campaigns that use influence maximization for harmful purposes (e.g., fake news, astroturfing)
- Modeling Uncertainty:
  - Incorporating bounded rationality and uncertainty in users' decision-making processes
  - Considering dynamic factors such as changes in public opinion or platform regulations

Conclusion:

- Influence maximization techniques are powerful but need to be used responsibly and ethically
- Balancing the benefits of rapid adoption with the potential for manipulation and harm

Theory vs. Practice in Social Influence Modeling

Challenges in Real-World Applications

In real-world applications of network influence, several challenges arise that make it difficult to implement theory-based models directly

Key challenges:

- Limited Data: In many cases, there is insufficient or incomplete data on network structures, individual behaviors, or interactions
- Dynamic Network Changes: Social networks are constantly evolving, with new nodes and connections forming, which complicates modeling and prediction
- Behavioral Variability: Individuals in networks often exhibit unpredictable or varying behaviors that are hard to model precisely

These challenges highlight the gap between theoretical models and practical applications of influence maximization in real-world scenarios

Limited Data on Network Topology

One of the major challenges in real-world network influence applications is the limited data on network topology. This can affect the accuracy of models and predictions

Key points:

- Incomplete Network Information: Social networks often lack full data on all nodes and their connections, limiting the ability to compute accurate centrality measures
- Data Privacy and Access: Access to detailed network data (such as personal interactions) may be restricted due to privacy concerns
- Sparse or Incomplete Relationships: Some connections may not be represented or may be weak, affecting the diffusion of influence

Despite these limitations, data from social media platforms or surveys can still provide valuable insights into network structure and influence

Dynamic Changes in Network Structure

Another challenge is the dynamic nature of social networks, where connections and behaviors change over time.
Key points:

- Evolving Connections: People's relationships and interactions evolve, making it difficult to predict how influence will spread in the future.
- Emerging Influencers: New nodes or influencers may emerge unexpectedly, changing the influence landscape.
- Time-sensitive Behaviors: Influence can have short-lived effects, meaning that strategies must adapt quickly to shifts in the network.

These dynamic factors complicate the prediction and control of influence spread, requiring real-time adaptation of strategies.

Behavioral Variability

Behavioral variability refers to the unpredictability of how individuals in a network behave, which adds complexity to influence models.
Key points:

- Unpredictable Adoptions: Individuals may adopt behaviors in unexpected ways, making it difficult to predict who will influence whom.
- Changing Preferences: People's preferences and behaviors change over time, influencing how likely they are to adopt new ideas.
- External Factors: Social, cultural, or environmental factors may impact the decision-making processes of individuals in the network.

This variability makes it challenging to create accurate models for predicting the spread of influence across a network.

Deeper Dive into Network Dynamics

A deeper understanding of network dynamics involves studying the evolving nature of networks and how this evolution influences the spread of influence. Key concepts in network dynamics include:

- Network Growth: How new nodes and edges are added to a network over time.
- Evolution of Relationships: How the strength and nature of relationships change over time, impacting information diffusion.
- Dynamic Centrality: How the centrality of nodes changes as the network evolves, which affects influence spread.

By modeling these dynamics, we can better understand how influence propagates over time and under different conditions.

Influence Models in Targeting Marketing Efforts

Influence models are widely used in marketing to target specific individuals or groups to maximize the impact of a campaign

These models predict how information spreads through a network and identify individuals who are most likely to amplify the marketing message

Key Concepts:

- Early Adopters: Influencers or individuals with high centrality who can quickly spread information to a large number of people
- Network Effect: The amplification of marketing efforts due to word-of-mouth or social sharing by key influencers

Seeding Strategies to Maximize Spread in Marketing Campaigns

Companies target individuals with high influence (seeds) in a social network to ensure that their campaigns are quickly disseminated across the network

Key Considerations:

- Number of Seeds: A balance between targeting enough influencers to initiate the spread and keeping costs low.
- Quality of Seeds: High centrality nodes are preferred for their ability to spread the message quickly and to many others
- Influence Maximization:
- Given a budget k, the goal is to select k nodes to maximize the expected spread of the campaign.

Using Influence Models to Understand Social Movements

Influence models help understand how social movements spread and gain momentum. These models can simulate how ideas, behaviors, or calls to action propagate through social networks

Key Elements:

- Core Activists: Individuals who initiate the movement (similar to seed nodes in marketing)
- Influential Spreaders: Highly connected individuals who can broadcast the message widely
- Thresholds: Individuals adopt the cause once enough of their peers (threshold) have adopted it

Influence models can predict how information will spread over time

- Spread: The total number of nodes influenced by the initial seed set.
- Speed: How quickly information spreads across the network.

In social media, influence models can predict how fast a news article or tweet will go viral based on network topology and the initial influencers

Predicting Virality Using Influence Models

Virality refers to the rapid spread of content across a network, where it is shared by a large number of individuals in a short time.

Influence models help predict which content is likely to go viral

Factors Influencing Virality:

- Initial Seed Set: Content shared by influential individuals is more likely to go viral
- Network Structure: Dense, well-connected networks facilitate faster spread
- Content Relevance: Content that resonates with a large portion of the network is more likely to be shared widely

Marketers and social media platforms use influence models to predict which posts, videos, or tweets are likely to go viral, allowing them to promote content more effectively

Empirical Studies on Weak and Strong Ties in Influence Models

Several studies have examined the role of weak and strong ties in influence propagation

These studies emphasize the complementary roles of weak and strong ties in spreading influence, with weak ties expanding the range of influence and strong ties deepening adoption within groups

- Granovetter (1973): Demonstrated that weak ties are crucial for spreading information across different social groups. Strong ties, while important for reinforcing influence, tend to form clusters that limit the spread of new ideas
- Aral and Van Alstyne (2011): Showed that weak ties provide access to novel information, making them essential for innovation and discovery in networks. Strong ties are better for rapid diffusion of ideas within close-knit communities but less effective for reaching new audiences
- Centola (2010): Found that while weak ties are essential for broadening the reach of influence, strong ties provide reinforcement that leads to higher rates of behavior adoption in social networks

Challenges in Translating Theoretical Models into Practical Systems

Although influence models provide valuable insights into network behavior, several challenges arise when translating theoretical models into practical systems:

- Complexity of Real-World Networks: Theoretical models often simplify network structures, assuming static or homogenous networks. In reality, networks are dynamic, heterogeneous, and influenced by external factors such as social context and evolving relationships
- Data Availability and Quality: Theoretical models require detailed and accurate data about the network. In practice, obtaining high-quality data on real-world social networks is difficult due to privacy concerns, incomplete datasets, and the sheer scale of modern social platforms
- Scalability: Many theoretical algorithms, especially those based on influence maximization, are computationally expensive and do not scale well to the size of networks such as Facebook or Twitter, which may consist of millions of users and billions of connections
- Human Behavior Complexity: Theoretical models often fail to capture the nuances of human behavior, including irrational decision-making, peer pressure, and emotional responses, all of which play a significant role in influence propagation

The Theory-Practice Gap

Several case studies demonstrate the challenges of applying theoretical influence models to real-world scenarios:

- Marketing Campaigns: Companies like Facebook and Twitter use influence maximization models to identify key users for advertising campaigns. However, the real-world performance of these models often falls short due to unpredictable user behavior, low engagement, and the difficulty in accurately predicting how content will be shared.
- Political Movements: During events like the Arab Spring, social network analysis helped identify key influencers. However, translating these theoretical insights into actionable strategies proved difficult due to the unpredictable nature of social movements, government interventions, and the diversity of actors involved.
- Health Campaigns: Influence models have been applied to encourage the spread of healthy behaviors (e.g., vaccinations). However, real-world campaigns face challenges like misinformation, cultural resistance, and logistical barriers not accounted for in theory

Recent Developments in Social Network Analysis

New Developments in Social Network Analysis

Social Network Analysis (SNA) have introduced new methods for understanding network structures and influence propagation:

- Dynamic Network Analysis: Modern SNA is moving beyond static models to account for dynamic networks where nodes and connections change over time. This is particularly relevant for social media platforms where user engagement and interactions fluctuate.
- Multilayer Networks: Models now analyze multilayer or multiplex networks, where individuals participate in different types of relationships (friendship, professional connections) that interact and influence each other
- Temporal Influence Models: With the rise of real-time social platforms, researchers are focusing on models capturing temporal aspects of influence, e.g., the speed of information diffusion and the impact of early adopters in fast-changing environments
- Behavioral Insights: Recent studies incorporate behavioral economics and psychology to better understand how real-world decision-making deviates from traditional rational agent models in network settings

Trends in Machine Learning for Network Influence

Machine learning (ML) is increasingly integrated with Social Network Analysis to enhance influence modeling and prediction:

- Graph Neural Networks (GNNs): GNNs represent a cutting-edge development in ML for networks, leveraging neural networks to learn complex patterns in graph structures. They are used for tasks such as node classification, link prediction, and influence estimation.
- Reinforcement Learning: RL is being applied to influence maximization problems, allowing systems to learn optimal seeding strategies through interactions with network. RL models are effective for adapting strategies in dynamic, evolving networks
- Deep Learning for Cascades: DL models are used to predict information cascades in social networks (e.g., viral content or rapid diffusion of ideas). These models capture long-range dependencies in the network and make predictions about influence propagation
- Transfer Learning: In networks where data is sparse, transfer learning allows influence models to be trained on similar networks and adapted to new contexts, reducing the amount of required training data

Ethics of Influence and Manipulation in Social Networks

The use of influence models in social networks raises important ethical considerations, particularly regarding the potential for manipulation:

- Manipulation and Control: Influence models can be used to manipulate users by targeting them with content/ideas designed to shape their behaviors and decisions, without their awareness/consent. This raises questions about autonomy and free will in digital spaces
- Behavioral Targeting: Social media platforms use influence models for targeted advertising and personalized content delivery, exploiting user data to predict and influence future behavior. This can border on manipulation when users are steered towards decisions that benefit advertisers or platform owners more than the users themselves
- Amplification of Bias: Influence models may reinforce existing biases in social networks, leading to echo chambers and filter bubbles, where users are exposed only to information that aligns with their existing beliefs. This can exacerbate societal polarization and misinformation

Ethical frameworks need to be developed to ensure that influence models are used responsibly, respecting user autonomy and preventing exploitation.

Case Studies on Ethical Issues in Social Media

Several high-profile case studies have highlighted ethical concerns in the application of influence models and manipulation in social media:

- Cambridge Analytica Scandal: used Facebook data to develop psychographic profiles of users and targeted them with personalized political ads during the 2016 U.S. presidential election and the Brexit campaign. This raised ethical concerns over privacy, manipulation, and the use of personal data without informed consent.
- Misinformation and Echo Chambers: Social media algorithms, designed to maximize engagement, have been criticized for promoting misinformation and creating echo chambers (users are repeatedly exposed to content that reinforces their existing views). This has been especially problematic in the context of health information (e.g., COVID-19 vaccine misinformation) and political polarization
- Dark Patterns in Social Networks: Platforms like Instagram and TikTok have been accused of using dark patterns, subtle design techniques that encourage users to spend more time on the platform or engage with content in ways that are not in their best interest, raising ethical concerns about user manipulation and addiction.

Innovations in Algorithms for Social Network Analysis

Recent innovations in algorithms for Social Network Analysis (SNA) have significantly improved our ability to understand and model influence in large-scale networks:

- Graph Neural Networks (GNNs):
- Random Walk-Based Algorithms:
- Hypergraph Analysis: Traditional social networks are modeled as simple graphs, but recent innovations involve using hypergraphs, where an edge can connect more than two nodes. This allows for more accurate representation of real-world relationships, such as group dynamics in social networks.
- Probabilistic Graph Models: Bayesian networks and Markov Random Fields are being integrated into SNA to deal with uncertainty and partial information, improving the robustness of models used in areas such as recommendation systems and disease propagation.

Technological Challenges in Scaling Influence Models

Scaling influence models to large networks presents several challenges:

- Computational Complexity: Influence maximization problems (e.g., selecting seed nodes) are NP-hard, meaning that the time to compute solutions increases exponentially with network size. Algorithms that scale to networks with millions of nodes remains challenging
- Memory Constraints: Processing massive social networks requires significant memory resources.
- Dynamic Networks: Many social networks are dynamic and evolving. Traditional influence models assume static networks, but scaling to handle time-varying networks introduces further computational overhead. Techniques such as incremental computation and streaming algorithms are being developed to address this.
- Data Privacy and Security: As influence models rely heavily on user data, scaling these models raises privacy /security concerns. Development of privacy-preserving algorithms is an ongoing challenge