On Studying Information Dissemination in Social-Physical Interdependent Networks

Mingkui Wei[1], Jie Wang[2], **Zhuo Lu[3]**, and Wenye Wang[2]

[1] Sam Houston State University

[2] North Carolina State University

[3] University of South Florida

Social Networks in Our Daily Life

- Online social networks is undergoing phenomenal growth in the recent decade, and is now the major medium for information dissemination.
  - Facebook: 2.27 billion monthly active users (2018 Oct)
  - Twitter: 330 million monthly active users (2019 Q1)
  - …

https://yourdigitalresource.com/social-media-marketing-in-detroit-musts/https-blogs-images-forbes-com-jaysondemers-files-2014-08-social-media-marketing/

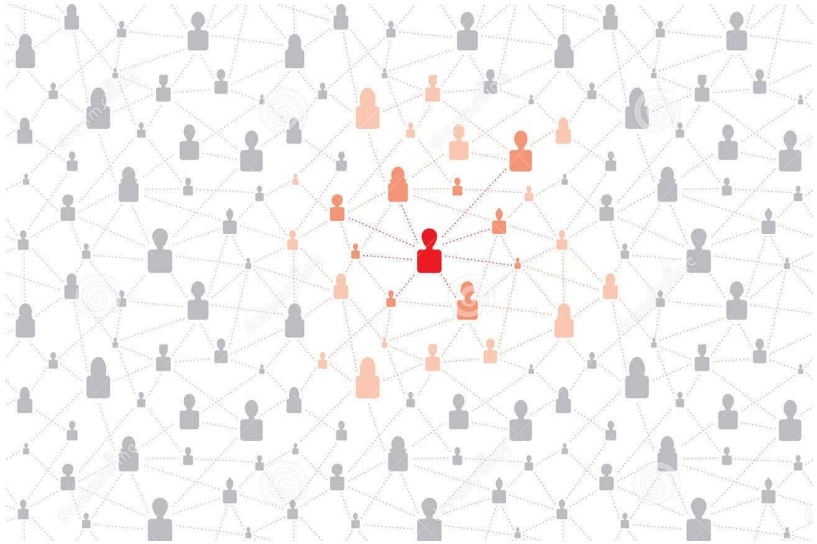
Information Dissemination in OSNs

- It is important to understand how information spreads among social networks
  - From users' perspective, it may result in more effective advertisement, campaign, public exposure, etc.
  - From regulator's perspective, it may helps in censorship, such as timely filtering and eliminating unlawful information and rumors.

https://www.istum.it/blog/?cat=7

Existing Works

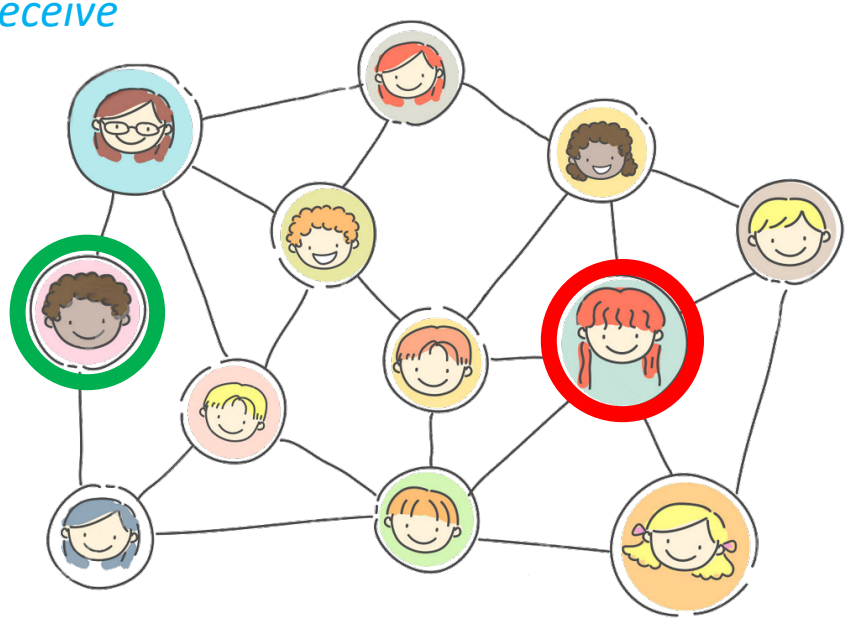
- Mainstream of study is based on graph theory and *epidemic models*.
  - Each user account is modeled as a *node*.
  - Each online relationship is modeled as an *edge*.
  - A piece of information is analogous to a virus that can *infect* a node.
  - A node can be either *susceptible* (S) to, *infected* (I) by, or *recovered* (R) from, the virus infection, and thus the SIR model.



https://www.pinterest.ru/pin/561401909776548615/

Existing Works

- Explicit/implicit assumption in epidemic models:
  - Virus infects node with certain probability.
  - The more adjacent nodes a node has, the more exposure it is to the virus, and the more likely it will be infected.
  - In other words, a *node's infection rate is positively correlated to its degree*.
- But, is it really the case in OSNs?
  - Is Alice more likely to *effectively receive* an information because she has 5 edges?



https://medium.freecodecamp.org/deep-dive-into-graph-traversals-227a90c6a261

Motivation

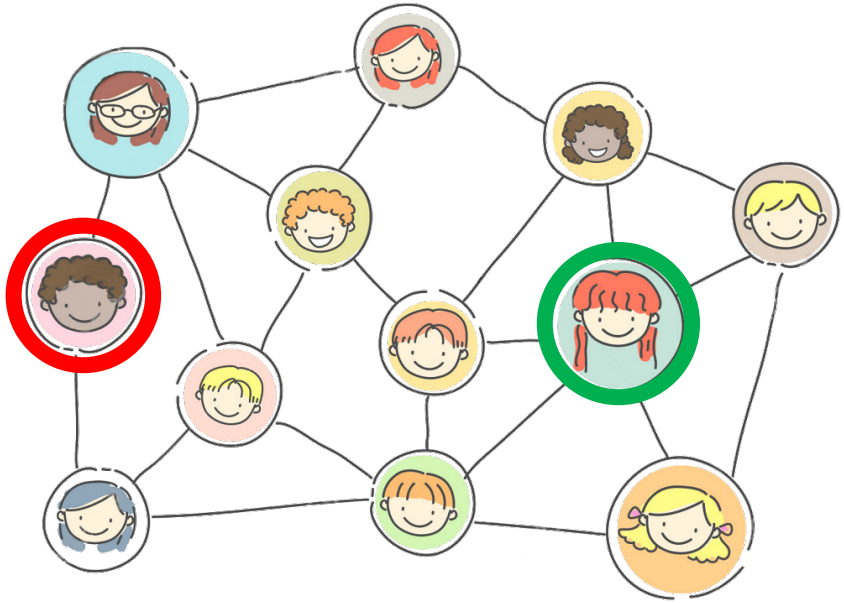
- Information overload -- the difficulty in understanding an issue and effectively making decisions when one has too much information about that issue (Wikipedia).
- How long are you willing to scroll up before you screw it up?
- How about when things multiply?

https://menawebagency.net/snowyish-no-social-day-wait/

https://giphy.com/gifs/scrolling-mH3aeWJbJ0JaM

Motivation

- We argues that from the perspective of a particular piece of information, the more degree a node has, the less likely it will effectively receive the information.
  - The infection rate is negatively correlated to a node's degree.
  - It is distinguish from epidemic models, and is specific for online social networks.



https://medium.freecodecamp.org/deep-dive-into-graph-traversals-227a90c6a261

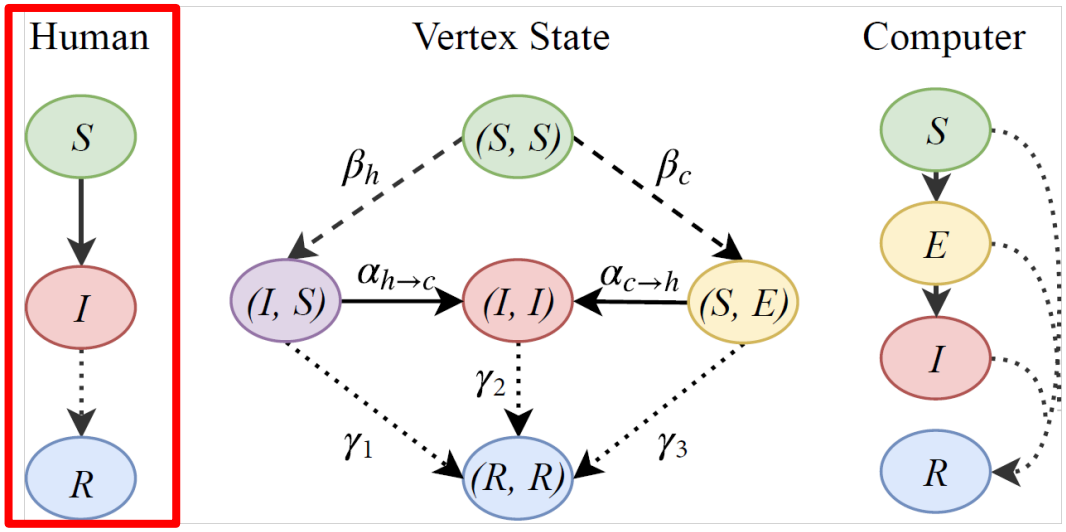
Social-Physical Interdependent (SPI) Model

- We model the social network as a dual-layered, asymmetric network.

- Dual-layer:
  - Human beings formulate an offline social network.
  - Their smartphones/computers formulate an online social network.

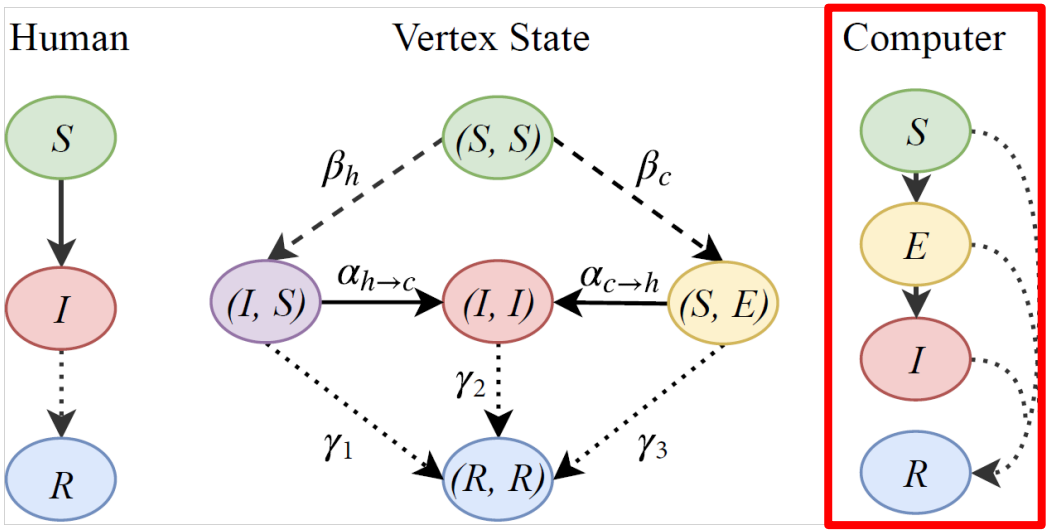
- Asymmetric:
  - Human controls his/her computer.
  - A computer can be “exposed” to an information, but won’t share it until the information is read, and decided to share, by its human owner.

Social-Physical Interdependent (SPI) Model



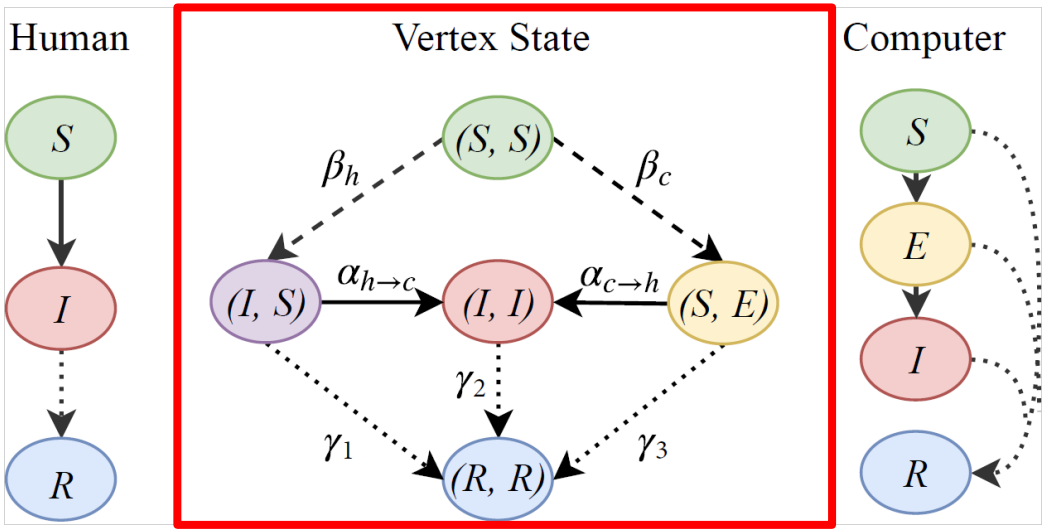
- Human nodes:
  - **Susceptible**: has not received the information either from offline nor from online network.
  - **Infected**: is aware of the information, and is about to share the information with certain probability.
  - **Recovered**: is either indifferent to share the information, or has shared and no longer interested to participate the sharing.

Social-Physical Interdependent (SPI) Model



- Computer nodes:
  - **Susceptible**: has not received the information either from online network.
  - **Exposed:** the information is arrived at the computer, but has not been read by the human user.
  - **Infected**: its human user is aware of the information, and is about to share the information with certain probability.
  - **Recovered**: its human user is either indifferent to share the information, or has shared and no longer interested to participate.

Social-Physical Interdependent (SPI) Model



- Model mechanism
  - A person get infected with probability β_h, while his computer get exposed with probability β_c.
  - With probability γ_1, an infected person will directly recover (i.e., the person is not interested in sharing).
  - With probability γ_3, an exposed computer will directly recover (i.e., the information is missed by its user).

Model Validation

- Objective: validate the proposed SPI model more accurately fits the information dissemination pattern in online social networks.
- Approach: case study via simulation.
- Dataset: unfortunately, we were not able to find a comprehensive dataset that contains accurate network connectivity, as well as information dissemination pattern.
- Tradeoff: we obtain partial information from different datasets, and validate the model by “matching the trends” (more on this).

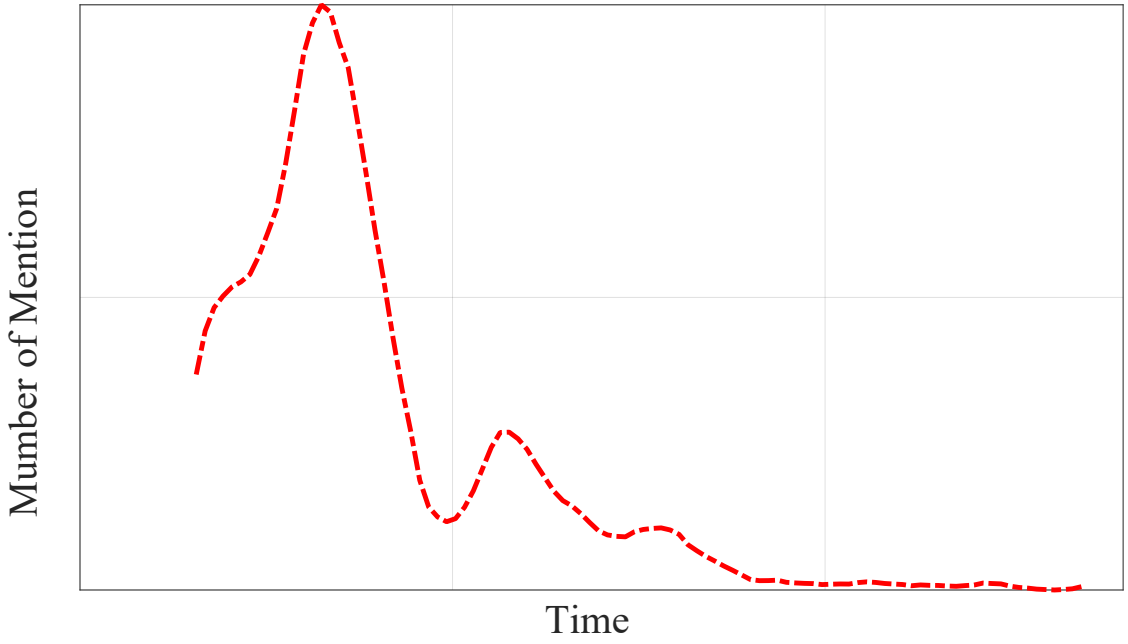
Simulation Setup

- Online social network graph: obtained by crawling www.epinions.com [20].
  - 75,879 nodes and 508,837 edges.

- Offline social network graph
  - Same nodes, but different topology.
  - Each node is randomly assigned a degree with the average as 6.

- Dataset
  - The popularity of a topic that was tracked by Memetracker [21].
  - The topic: March 20th 2009, President Obama joked about his bowling skills, saying “It was like a Special Olympics, or something” on The Tonight Show[2]. Considered offending to certain populations, this news got popular in the next following days.

Simulation Setup

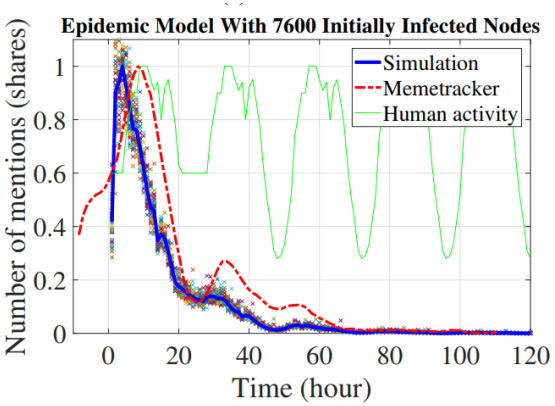


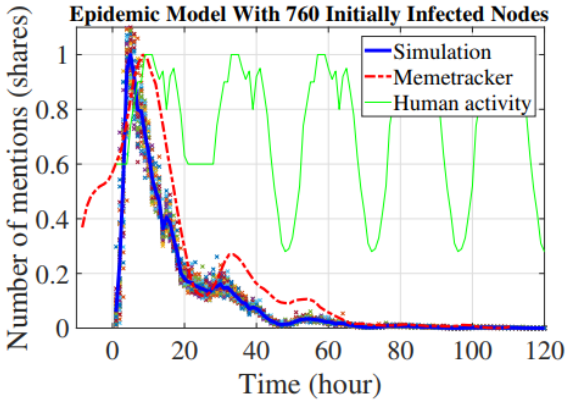
- Time-wise variation of the simulated dateset.
  - X-axis is time.
  - Y-axis is the number of the topic being mentioned online.
- Because the dataset is not obtained from the simulated social network, we seek for “pattern match” rather than numerical match.

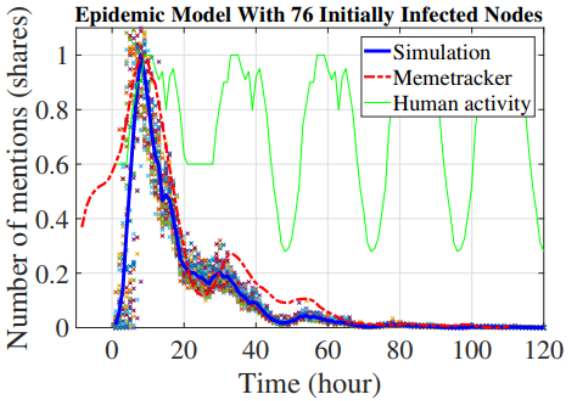
Simulation Setup

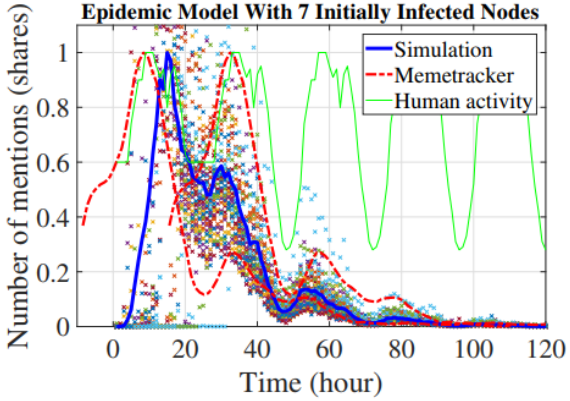
- We compare the simulation result of the existing epidemic model, and the newly proposed SPI mode.
- For each model, we run multiple simulations. And in each simulation, we adjust above-mentioned parameters, with the objective to best match the simulation result to the real world dataset.

Simulation Result for Epidemic Model



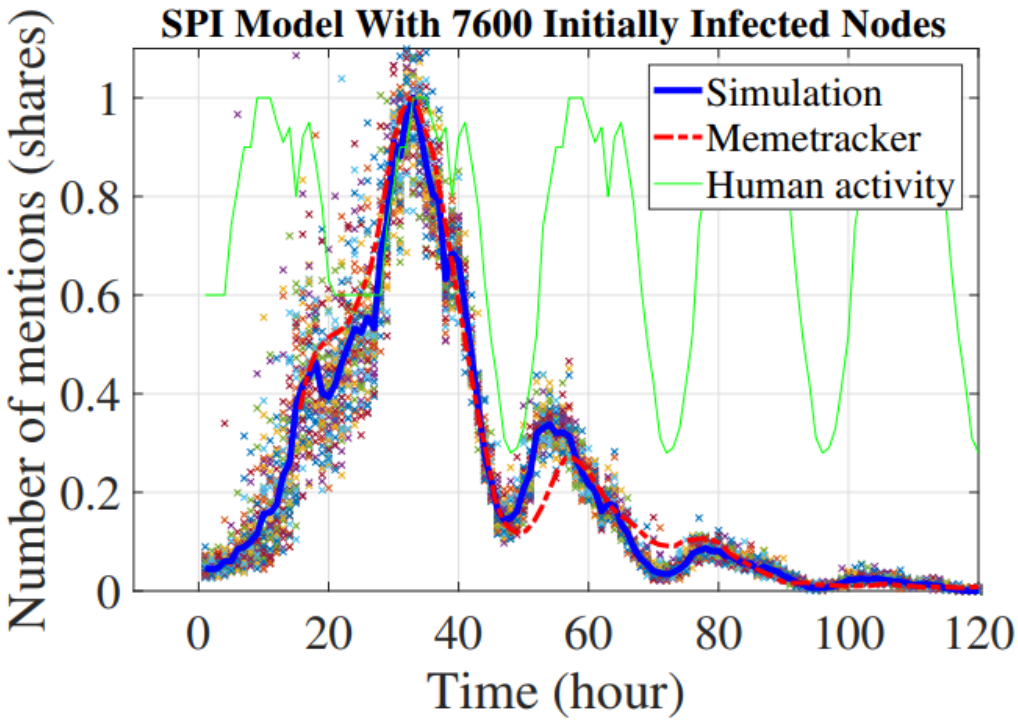






- 4 simulations with different initial number of infected nodes.
- Remind the objective is to match the blue line to the red line.
- Epidemic model isn't able to catch the "slow start" initial phase, nor the 2nd and 3rd wave.

Simulation Result for SPI Model



- Epidemic model presents exponential increase at the beginning, due to the assumption that degree and infection rate are positively correlated.
- SPI model assumes the inverse, and thus has a much slower initial phase, as well as slower diminishing rate.
- It can be observed that SPI model more accurately follows the real world dataset.

Conclusion

- We observe that the epidemic model is not an accurate reflection of how social network behaves.
- Based on this observation, we proposed the Social-Physical Interdependent (SPI) model, in which it is assumed that a node's infection rate is negatively correlated to its degree.
- We validated the SPI model with real world dataset, and demonstrate that compared to epidemic model, the SPI models follows the dataset more accurately.
- Our future work includes derive theoretical analysis to more accurately depict the SPI model.

Thank you!