PART IV:
DIFFUSION ANALYSIS

When exploring the question of how online social networks affect the spread of ideas, innovation, and behaviour, the first key aspect to consider is a network's structure or topology. Who a person is connected to is who they come into contact with, and therefore who they "allow" to have an impact on their behaviour. The extent to which this affects the individual is predicated on how they "choose" to come into contact with this person: how close they are, how frequent the contact is, is it through a close knit group of mutual friends or through separate interactions? Given the many factors that determine the diffusion process, a common first step approach is to examine the network for "small world" features. Small worlds characterise a fascinating aspect of large social networks that find that even if nodes in a network are not directly connected, the likelihood that their neighbours are each others neighbours is high. This means any given node in a network can be reached in a small number of steps, which translates to a short average path length and a high clustering coefficient. Naturally, this creates an easy path for diffusion.

Though exhibiting small world properties, most real world networks have scale-free degree distributions where some nodes attract links or connections at a higher propensity than others. This creates a complexity in the diffusion process because it means people display heterogeneous degrees of connectivity and therefore vary in their vulnerability to social influence mechanisms. In other words, while one person can adopt a behaviour after coming in contact with a single "infected" person, another person requires multiple sources of contact with a group of "infected" persons before they can be convinced of adopting a behavior. A common hypothesis of the role communities play in diffusion problems is that the presence of redundant (path independent) ties, found in highly clustered communities or networks, can act as a social reinforcement mechanism (or multiplier effect). First, by allowing infections to easily spread within communities, then, by penetrating outside communities where users are more likely to display higher levels of resistance. Contrarily, if not enough paths or "long ties" that connect users between communities exist, the presence of redundant (path independent) ties can also act as a social containment mechanism (or trap) that fails to permeate outside communities.

Methods

There are various approaches to addressing the diffusion of ideas or innovation in complex contagion problems. One way is to model them as either passive diffusion, active diffusion, or mixed diffusion.

> **What is Active, Passive and Mixed Diffusion**
> **Passive** diffusion assumes diffusion relies only on peer pressure; that the more an individual is exposed to a given content the more likely it will adopt it. Like the spread of a disease, adoption happens once a certain degree of exposure is reached, independent of the individuals willingness or personal interest. **Active** diffusion assumes that the diffusion process purely relies on individual choice and preference; that each node can choose to adopt once they come into contact with the information that serves their interests. This ignores any social influence from peers. Whereas **Mixed** diffusion combines both passive and active diffusion processes and assumes that individual's interests act as a preferential schema for adoption, and at the same time, does not neglect the role of peer-pressure mechanisms. Therefore the "infection" or decision to adopt is evaluated by individual preference once a sufficient peer pressure (exposition) is achieved.
>
> The **profile-threshold model** is designed to support mixed behaviors and simulate an active and passive adoption process. In this model, a susceptible node has two distinct and mutually exclusive behavioral alternatives: to adopt based on peer-pressure and self-interest (infected) or to not adopt (not infected). Once exposed to a certain amount of peer pressure, adoption occurs when personal preferences and "infection" content coincide. The model works as follows:
>
> 1. Each node starts with its own profile (γ), threshold (τ), and status (infected or susceptible). Nodes infected at time *t=0* are called initial seeds;
> 2. During each iteration time *t*, every node is observed: if the percentage of node *i*'s neighbours that were infected before time *t* is greater than its threshold (τ), then a random number is drawn, and if that random number is above node *i*'s profile (γ), node *i* becomes infected.

I simulate multiple diffusion processes using the **profile threshold model** and compare outcomes from specific positions set as initial seeds of infection (first adopters). The simulation mechanisms of the 4 positional infection models plus random infection model are presented below.

- **Random Infection**: I evaluate the diffusion process of randomly selected nodes as initial seeds and assess their capacity as a baseline or counterfactual.

- **Leader Infection**: I evaluate the diffusion process of *Leader* types as initial seeds and assess their capacity as central hubs of authority, capable of leading a movement and gaining the respect of their peers.

- **Activator Infection**: I evaluate the diffusion process of *Activator* types as initial seeds and assess their capacity as outspoken voices of the movement, capable of inciting activism and support.

- **Spreader Infection**: I evaluate the diffusion process of *Spreader* types as initial seeds and their capacity as bridge links, capable of reaching distant and isolated users.

- **Observer Infection**: I evaluate the diffusion process of *Observer* types as initial seeds in their capacity as uninvolved outsiders, capable of bypassing the weight of social pressure and popularism.

Steps in this module are as follows:

1. **Community Diffusion** -> to compare the diffusion process within community and outside community of each position set as initial seeds in the network.

2. **Positional Diffusion** -> to trace the diffusion trend during peak discourse of each position set as initial

seeds in the network.

1 COMMUNITY DIFFUSION

When predicting the success of a *meme*, hashtag, or protest movement in online social networks, two main predictions are made on how communities can affect diffusion processes. They can either promote diffusion by reinforcing the behaviour, or they can inhibit diffusion by trapping the behaviour. The first assumes a social contagion process where there exists multiple paths or opportunities for exposure, while the second assumes a closed containment process where members of outside populations are protected from the exposure. To take it a step further, I ask how this impact differs when the diffusion process is introduced by specific users occupying certain roles or positions in the network.

1.1 Cohesion in Community

One key factor determining the diffusion outcome is the cohesion or "small world" aspect of the communities containing the initial seeds. If the infection emanates from communities with a high incident of clustering, then we can expect many redundant ties or a social reinforcement mechanism. If the infection is started in communities with high average path lengths, then we can expect a slower or contained diffusion process. To estimate this, I first determine the communities containing each initial seed and measure the average path length and average clustering coefficient for each communitity within each role. I plot these measures using a bubble plot figure, where each bubble size is scaled to reflect the number of communities represented by each role.

1.2 Diffusion In and Out of Community

The relationship between clustering and diffusion is observed by measuring the extent of the spread within community and the extent of the spread outside community for each role. A high occurrence of spread within community relative to outside community is consitent with complex contagion where redundant ties are leveraged to achieve greater virality. Furthermore, to assess whether early stages of diffusion can predict how a discourse will spread, I estimate models for both early stage diffusion, which I define as the accumulated infection rate from the very first day (first tweet reply) to the day before the peak of the discourse (i.e. the days leading up to the the discourse trending), and late stage diffusion, which I define as the total infection rate from the very first day (first tweet reply) to the very last day (last tweet reply) of the observation period.

2 POSITIONAL DIFFUSION

Finally, as an alternative model and robustness check, I simulate the rate and extent of possible diffusion in peak discourse by assigning specific users from each role as initial seeds and tracing their diffusion trend using the **profile threshold model**. In addition to the diffusion trend, I visualise a snapshot graph showing the infection rate for each simulation.

3 REFERENCES

Goyal, Sanjeev., Connections: An Introduction to the Economics of Networks, Princeton, NJ: Princeton University Press, 2007.

Jackson, Matthew O., Social and Economic Networks, Princeton, NJ: Princeton University Press, 2008.

Wasserman, Stanley and Katherine Faust, Social Network Analysis, New York, NY: Cambridge University Press, 2007.

Watts, Duncan J., Small Worlds: the dynamics of networks between order and randomness, Princeton, NJ: Princeton University Press, 2005.

Milli, L., Rossetti, G., Pedreschi, D. and Giannotti, F., "Active and passive diffusion processes in complex networks,"Applied Network Science, 2018, 3 (42).

Centola, Damon., "The Spread of Behavior in an Online Social Network Experiment,"Science, 2010, 329.

Valente, Thomas W., "Network Interventions,"Science, 2012, 337.

Moody, Jim., "Diffusion and Peer-Influence," Social Networks and Health Workshop, 2016. Available at: https://sites.duke.edu/dnac/04-diffusion-and-peer-influence-2016/