

MATHEMATICAL INSTITUTE

NETWORKS C5.4, HT 2019

MSC SPECIAL TOPIC

---

# Complex contagion patterns in small-world networks

---

*Candidate Number:*

*Lecturer:*

Prof. Renaud LAMBIOTTE

May 20, 2019



# 1 Introduction

Over the past decades, social scientists have been interested in how the social networks structures affect the dynamics of diffusion. In the late 1960s, Mark Granovetter interviewed people who had recently changed employers to understand how they heard about their new jobs. He realised that the interviewees became aware of the new job opportunity more often via acquaintances than via close friends. In 1973, he proposed the famous "strength of weak ties" hypothesis: in small worlds, "whatever is to be diffused can reach a larger number of people, and traverse a greater social distance, when passed through weak ties (e.g acquaintances) rather than strong ones (e.g close friends)" [1]. Although this theory applies to simple contagions (information or disease transmission), Centola et al (2007) demonstrated more recently that the strength of weak ties does not generalize to complex contagions (urban legends, social norms, etc.). More particularly, they show that the ability of long-range ties to accelerate the diffusion process highly relies on the network structure. Therefore, while long-range ties may facilitate simple contagions, they can have a harmful effect for spreads requiring social reinforcement from multiple contacts [2]. In this paper, we use Facebook data to highlight the ties that sustain complex contagion processes within small-world networks. We also investigate the nature of the nodes that are decisive in these processes. We found that complex contagions require a relatively small critical mass of infected nodes above which most nodes are ensured to be contaminated. Along with this, we found that the chances of reaching the critical mass are significantly higher when initiating the process at a node with high degree and high second-degree network size. Lastly, we found that although the complex contagions may mostly benefit from weak ties, there exists network configuration for which they take more advantage of the ties of intermediate strength. Section 2 specifies the framework of the study. Section 3 provides the theoretical background of the experiments. Section 4 shows that the Facebook networks used in our study are small-world networks. Section 5 presents the experiment methods and their results.

## 2 Framework

### 2.1 Context

In social networks, complex contagion is a contagion for which multiple sources of exposure to a behaviour are required before an individual adopts this behaviour.

It differs from simple contagion for which one contact with an infected neighbour is enough to become infected. Typical complex contagions are phenomena that are risky or lack of credibility and/or legitimacy: urban legends, social norms, the willingness to participate in collective actions like strikes etc. Conversely, simple contagion phenomena like communicable diseases and common information do not require social reinforcement.

For the sake of simplicity, we refer to any agent’s behavioural change as an ”infection” in the remainder of the paper.

## 2.2 Data

Our research aims at highlighting key contagion patterns in complex contagion by using Facebook data. The *Stanford Network Analysis Project* (SNAP) [15] has published anonymized Facebook ego-networks (friends lists). Ego networks consist of a focal node (”ego”) and the nodes to whom ego is directly connected to (these are called ”friends”) plus the ties among its friends. We use 7 of these real-world social networks. For the sake of our study, we remove the ego along with its incident edges. In order to perform contagion processes, we restrict our work to the largest connected component of each of them. Figure 1 gathers the main characteristics of each resulting network. We observe that the networks and their largest connected component have nearly the same size.

|                             |                        | EGO NETWORKS |        |       |        |        |       |      |
|-----------------------------|------------------------|--------------|--------|-------|--------|--------|-------|------|
|                             |                        | A            | B      | C     | D      | E      | F     | G    |
| Whole network               | Nodes                  | 333          | 1,034  | 224   | 786    | 747    | 534   | 52   |
|                             | Edges                  | 2,519        | 26,749 | 3,192 | 14,024 | 30,025 | 4,813 | 146  |
| Largest connected component | Nodes                  | 324          | 1,034  | 224   | 775    | 744    | 532   | 44   |
|                             | Edges                  | 2,514        | 26,749 | 3,192 | 14,006 | 30,023 | 4,812 | 138  |
|                             | Clustering coefficient | 0.52         | 0.53   | 0.54  | 0.47   | 0.64   | 0.55  | 0.46 |
|                             | Average degree         | 16           | 52     | 29    | 36     | 81     | 18    | 6    |
|                             | Average path length    | 4            | 3      | 3     | 3      | 3      | 3     | 3    |
|                             | Diameter               | 11           | 9      | 9     | 9      | 7      | 9     | 5    |

Figure 1: Descriptive statistics

Figure 2 depicts the network F. We notice that (i) small groups are densely connected: there are many short ties within ”small” groups of nodes. However, (ii) the network is

sparsely connected at the larger scale, with few ties between these dense and "small" groups of nodes. These two properties are key characteristics of small-world networks and will be discussed theoretically in Section 4.

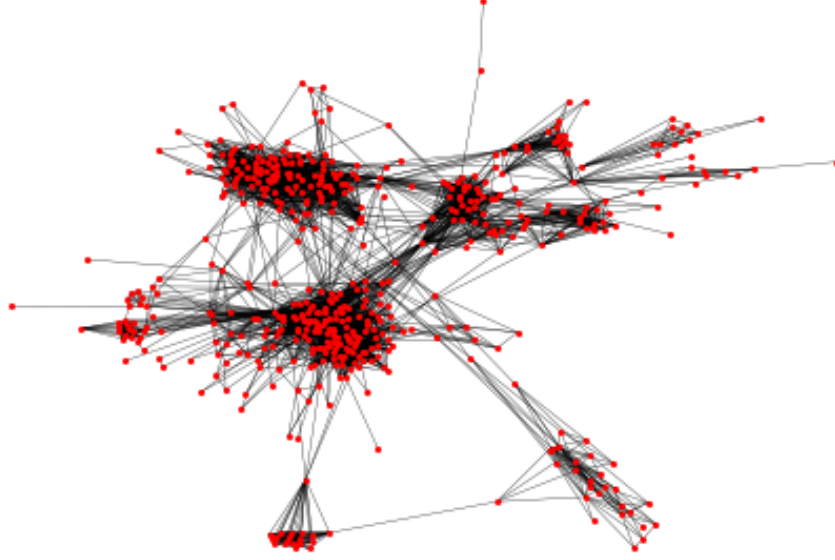


Figure 2: Network F

### 3 Theoretical background

This section presents the theoretical tools used in the experiments. Section 3.1 defines contagion theoretically. In section 3.2, we define indicators accounting for each node's social capital. In section 3.3, we specify two indicators quantifying the strength and length of the ties of the networks.

#### 3.1 Threshold and contagion rate

As in [2] and [8], we employ a common model of contagion: the relative threshold model, where a uninfected node becomes infected only if a critical fraction of its neighbors has become infected. The agents of the networks can change their state only from uninfected to infected in a deterministic fashion:

**Definition 3.1.** The relative threshold  $\alpha \in (0, 1)$  is the minimal fraction of infected neighbors required so a node becomes infected. The number of infected neighbors required for a node with  $n$  neighbours is  $z = \alpha * n$ .

**Proposition 3.1.** A contagion is complex if  $z \geq \frac{1}{n}$  and simple if  $z = \frac{1}{n}$ .

At each time step of the process, each agent reassesses her/his neighborhood. Assume  $\alpha = 0.5$  for instance. Then if an agent is uninfected but has more than 50% of infected neighbours, s/he becomes infected. If an agent is infected, s/he stays infected up to the end of the process.

Initially, a small set of nodes  $A$  (the innovators that initiate a movement...) have threshold  $z = 0$ . Starting with  $A$ , the contagion propagates throughout the population until no more agents can be infected. This happens when all agents are infected or when no uninfected agent has at least  $z = \alpha * n$  infected neighbours.

**Definition 3.2.** The infection reach of a contagion is the proportion of nodes infected at the end of the process.

### 3.2 Node properties

In economic terms, capital is a store of wealth and assets; social capital is the store of behaviours and norms in any large group that lets its members support one another. Here, we try to quantify the social capital of each agent (node). The literature distinguishes the bonding and the bridging capitals [7]:

- The bonding capital is an increase in the depth of connections and trust within a relatively homogeneous group (same ethnicity, language, religion, educational background...). It often includes ties with family and extended kin.
- The bridging capital is an increase in connections among relatively heterogeneous groups. They typically occur through institutions (schools, clubs, workplaces) and connect people from different backgrounds.

To measure nodes' social capital, we introduce the same quantitative proxies as Metaxa-Kakavouli et al (2018) [11]. In the remainder, we note  $N(v_i)$  the neighbourhood of node  $v_i$ :

**Definition 3.3.** The neighbourhood of a node  $v_i$ , called  $N(v_i)$  is the set of  $k_i$  nodes  $v_{i1}, v_{i2}, \dots, v_{ik_i}$  with which it shares an edge. We call  $k_i$  the degree of node  $v_i$ .

### 3.2.1 Bonding capital

We define two metrics of bonding capital: the first-degree network size and the local clustering coefficient.

**Definition 3.4.** The first-degree network size of a node is its degree.

**Definition 3.5.** The local clustering coefficient (of node  $v_i$ ) is defined by:

$$C_i = \frac{2T_i}{k_i(k_i - 1)} \quad (1)$$

where

- $k_i$  is the degree of node  $v_i$
- $T_i$  is the number of triangles (mutually connected three nodes) including  $v_i$

Equivalently, we note that  $\frac{k_i(k_i-1)}{2} = \binom{k_i}{2}$  is the number of possible connections in  $N(v_i)$  (between  $v_i$ 's neighbours) and  $T_i$  is the number of actual connections in  $N(v_i)$ . The clustering coefficient then quantifies the proportion of neighbours that are neighbours with each other (Figure 3). This metric makes sense to account for bonding capital as people with many connected friends are more likely to be embedded in a network of strong, close relational ties.

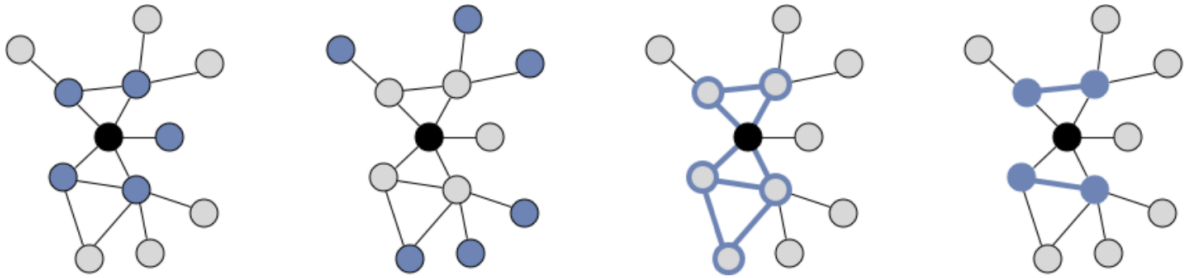


Figure 3: These graphs illustrate a small-scale example of a social network. From left to right: (1) highlights in blue the person's first degree network (their number of friends); (2) highlights second-degree network size (friends of friends); (3) highlights two bi-connected components (two interconnected groups of friends which the person is part of); (4) highlights all pairs of the person's friends who are friends with each other, from which clustering coefficient is computed. From [11].

### 3.2.2 Bridging capital

We define two metrics for bridging capital: the second-degree network size and the span.

**Definition 3.6.** The second-degree network size of a node  $v_i$  is its number of unique friends of friends: it is the number of unique nodes in  $\bigcup_{v_{ik} \in N(v_i)} N(v_{ik})$ .

Altenburger et al (2017) suggest that variety among friends-of-friends can provide fundamentally different and valuable information about people’s social networks [12]. This measure then examines heterogeneity in a person’s extended network and is a proxy for bridging capital, which relies on weaker, more diverse ties (Figure 3).

**Definition 3.7.** Biconnected components are maximal subgraphs such that the removal of a node (and all incident edges) will not disconnect the subgraph.

As shown in Figure 3, nodes may be part of more than one biconnected component. Those nodes are articulation points:

**Definition 3.8.** An articulation point is a node whose removal (along with all its incident edges) increases the number of connected components of a graph [13].

**Definition 3.9.** The span of a node is the number of biconnected components it is part of.

The only nodes with a span strictly above 1 are the articulation points. The node span was introduced as a measure of the cohesion of a network, since “biconnected components allow information to flow freely throughout the community” [14]. This kind of wider group cohesion that relies on intermediaries aligns with bridging capital too.

## 3.3 Strength and length of edges

### 3.3.1 Strength

In Granovetter’s usage, “tie strength” has a double meaning: relational and structural. The relational meaning refers to the strength of the tie as information: weak ties connect acquaintances, while strong ties connect close friends whose interactions are frequent and/or affectively charged. In this sense, strong ties increase the trust and the exposure we incur from contagious intimates. The structural strength of a tie refers to the ability of a tie to facilitate diffusion, cohesion, and integration of a social

network by linking otherwise distant nodes. Granovetter suggests that ties that are weak in the relational sense are often structurally strong in simple contagion. As a matter of fact, although they are weak information conduits, casual friendships are more likely to be formed between socially distant agents with few common friends. Therefore, they provide access to new information and greatly increase the rate at which information propagates. However, Centola et al. (2007) demonstrates that Granovetter’s strength of weak ties finding does not generalize to complex contagions [2]. Indeed, for the spread of information, ties between socially close agents tend to be redundant: ”in my group of 3 close friends, if one of my two friends tells me something, it is likely that my other friend already knows it too and therefore I do not need to tell to her/him”. However, when behaviour activation requires confirmation or reinforcement from two or more sources, this redundancy becomes an essential pathway for diffusion: short ties (between close friends) are therefore structurally strong too. In this context, we want to understand the correlation between the structural and relational strength: do complex contagion benefit from weak or strong ties? We try to answer this question in our experiments. Following Onnela et al (2007), we define the relational strength of a tie by the proportion of common friends of its two end nodes [9].

**Definition 3.10.** The relative neighborhood overlap of the edge linking two nodes  $v_i$  and  $v_j$  is given by:

$$O_{ij} = \frac{n_{ij}}{(k_i - 1) + (k_j - 1) - n_{ij}} \quad (2)$$

where

- $k_i$  is the degree of node  $v_i$
- $n_{ij}$  is the number of common friends of  $v_i$  and  $v_j$  ( $\text{card}(N(v_i) \cap N(v_j))$ )

We note that  $O_{ij} = 0$  if  $v_i$  and  $v_j$  have no common friends and/or acquaintances; conversely,  $O_{ij} = 1$  if they have the same friends and/or acquaintances. This is in keeping with the idea that the stronger a relationship between two people, the more their friends are likely to overlap.

### 3.3.2 Length

As said previously, edges do not generally have the same structural strength depending on whether they connect socially-close or distant nodes. To quantify edge lengths, we provide a few definitions taken from [10].



**Definition 3.11.** An edge joining two nodes  $v_i$  and  $v_j$  is a global bridge if deleting the edge would cause  $v_i$  and  $v_j$  to lie in two different components.

**Definition 3.12.** An edge joining two nodes  $v_i$  and  $v_j$  is a local bridge if its endpoints  $v_i$  and  $v_j$  have no friends in common.

**Proposition 3.2.** An edge is a local bridge if and only if deleting it would increase the distance between its endpoints to a value strictly more than two.

*Proof.* If we removing the edge between  $v_i$  and  $v_j$ , the distance between them is at least 2. If it is exactly 2, it means that they have a common friend through whom the infection propagates. If the distance is strictly above 2, the two nodes do not have a common neighbor and the edge is a local bridge.  $\square$

**Proposition 3.3.** The span of a local bridge is the distance its endpoints would be from each other if the edge were deleted.

With this in mind, we define the length of an edge as the distance its endpoints would be from each other if the edge were deleted:

- The length of a local bridge is its span.
- The length of a global bridge is  $\infty$  (in the experiments, the length of global bridges is set to 10.)
- The length of any other edge is 2.

We note that the bridges are the only edges with a relative neighbourhood overlap equal to 0.

Figure 4 displays the correlation coefficients between the length and strength of the ties. The correlation coefficients are all negative: the stronger the tie, the shorter. This intuitively suggests that people with many common friends have a close relationship.

|          |       |
|----------|-------|
| <b>A</b> | -0.15 |
| <b>B</b> | -0.06 |
| <b>C</b> | -0.11 |
| <b>D</b> | -0.05 |
| <b>E</b> | -0.02 |
| <b>F</b> | -0.08 |
| <b>G</b> | -0.32 |

Figure 4: Correlation coefficients between strength and length of ties

## 4 Small-world hypothesis

Following Milgram's small world experiment [3] published in 1967, Watts and Strogatz (1998) have defined a small-world network as a network "highly clustered and with a small characteristic path length" [4]. In other words, small-world networks (i) contain sub-networks which have connections between almost any couple of nodes within them, and (ii) most pair of nodes are connected by a short path.

Recall that an E-R random graph, say  $G(N, p)$ , is constructed by connecting each pair of the  $N$  nodes with probability  $p$  [6]. E-R graphs are low clustered and have a short average shortest path length. For this reason, it is relevant to compare networks to an (equivalent) E-R graph with the same average degree. Humpries and al. (2008) propose a theoretical equivalent definition of small-networks along with a quantitative characterisation [5]:

**Definition 4.1.** A network with  $N$  nodes and  $m$  edges is a small-world network if it has a similar path length but greater clustering of nodes than an equivalent Erdős-Rényi (E-R) random graph with the same  $N$  and  $m$ .

**Proposition 4.1.** Let  $G$  be a network. Let  $L_G$  be the mean shortest path length of  $G$  and  $C_G$  its clustering coefficient. Let  $L_{rand}$  and  $C_{rand}$  be the corresponding quantities for the equivalent E-R random graph. Then  $G$  is a small-world network when  $\frac{L_G}{L_{rand}} \geq 1$  (length ratio) and  $\frac{C_G}{C_{rand}} \gg 1$  (clustering ratio).

We test the small world hypothesis on the 7 networks. For a given network  $G$  with  $N$  nodes and an average degree  $d$ , we create an equivalent E-R random graph  $G_{ER}(N, p)$  with  $p = \frac{d}{N-1}$ . Indeed, recall that the average degree of  $G_{ER}(N, p)$  is  $(N-1)p$  which is equal to  $d$ . We compute the consequent distance and clustering ratios. We renew this process (creating E-R random graph) 1000 times for each network. Resulting average ratios are shown in Figure 5. We observe that the 7 networks are small worlds.

| Network Name | Clustering ratio | Length ratio |
|--------------|------------------|--------------|
| <b>A</b>     | 10.79            | 1.56         |
| <b>B</b>     | 10.53            | 1.46         |
| <b>C</b>     | 4.24             | 1.33         |
| <b>D</b>     | 10.07            | 1.43         |
| <b>E</b>     | 5.88             | 1.35         |
| <b>F</b>     | 16.07            | 1.38         |
| <b>G</b>     | 3.19             | 1.15         |

Figure 5: All the 7 networks satisfy the small-world networks conditions.

Another property of small-world networks is the presence of nodes with high degrees, called hubs. These small-world properties are found in many real-world phenomena. As a matter of fact, many real-world networks are believed to have degree distributions with heavy tails. This is the case of the networks B, C and F for instance (Figure 6). Note that although these networks have very different degree distributions, they are all small worlds.

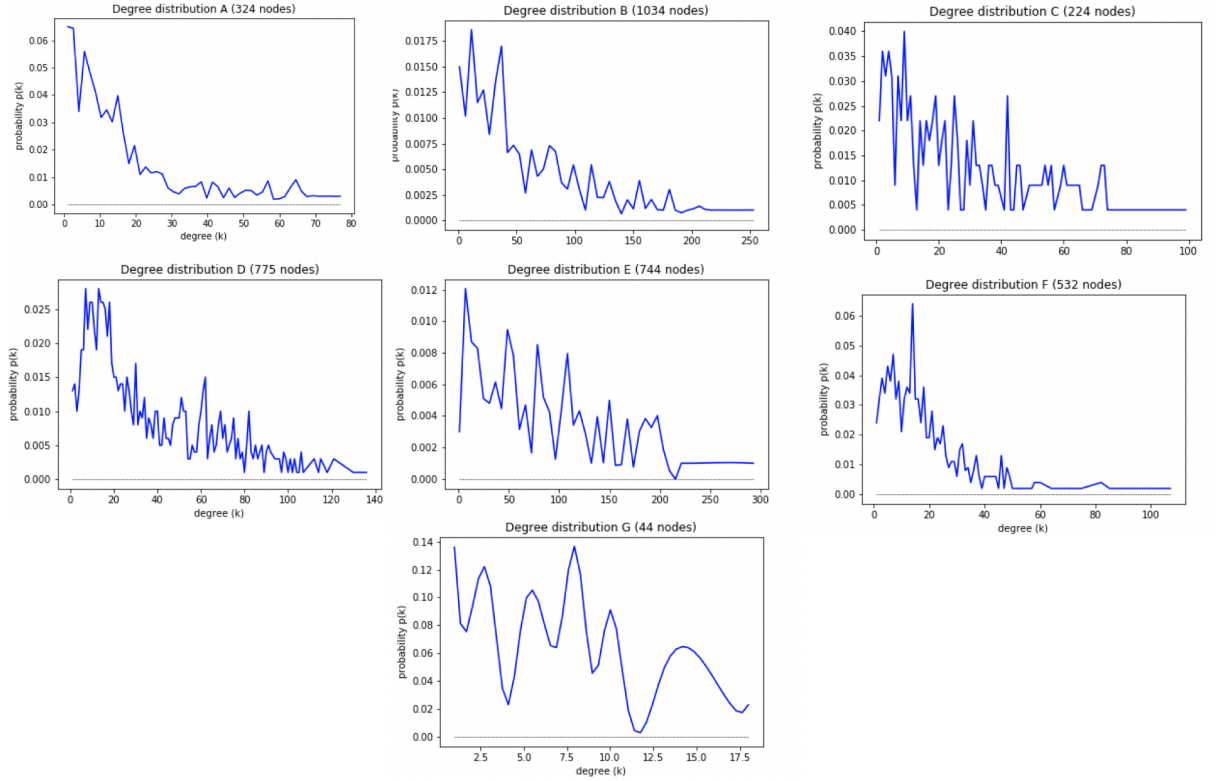


Figure 6: Degree distribution

In *Here Comes Everybody: The Power of Organizing Without Organizations* (2008), Clay Shirky discusses the implications of small world properties [7]. At the small scale, groups are tightly connected (high clustering coefficient) so they operate as both amplifiers (the more your friends care about a piece of information, the likelier you are to hear about it) and filters (things that none of your friends or their friends care about are unlikely to get to you) of information. At a larger scale, the highly connected individuals (hubs) connects sparsely the small dense groups, therefore holding the overall structure together. They also allow information to propagate efficiently as they dramatically lower the path lengths between any two nodes.

## 5 Experiments

In their research on perturbed regular lattice (usual theoretical representation of small-worlds), Barash et al (2012) discovered "the presence of a critical mass of infected nodes [...] prior to which contagions are fragile and are highly dependent on the idiosyncrasies of local network structure" [8]. Our research first investigates the

validity of these findings on Facebook small-world networks. It then focuses on understanding what characteristics differentiate the cascades that do not reach the critical mass and die out early from those which reach it and infect the majority of the networks. Section 4.1. highlights the presence of a critical mass in the networks. Section 4.2 investigates which nodes the contagion takes profit from to reach the critical mass. Section 4.3 discusses which edges are determinant for powerful contagions.

## 5.1 Critical mass

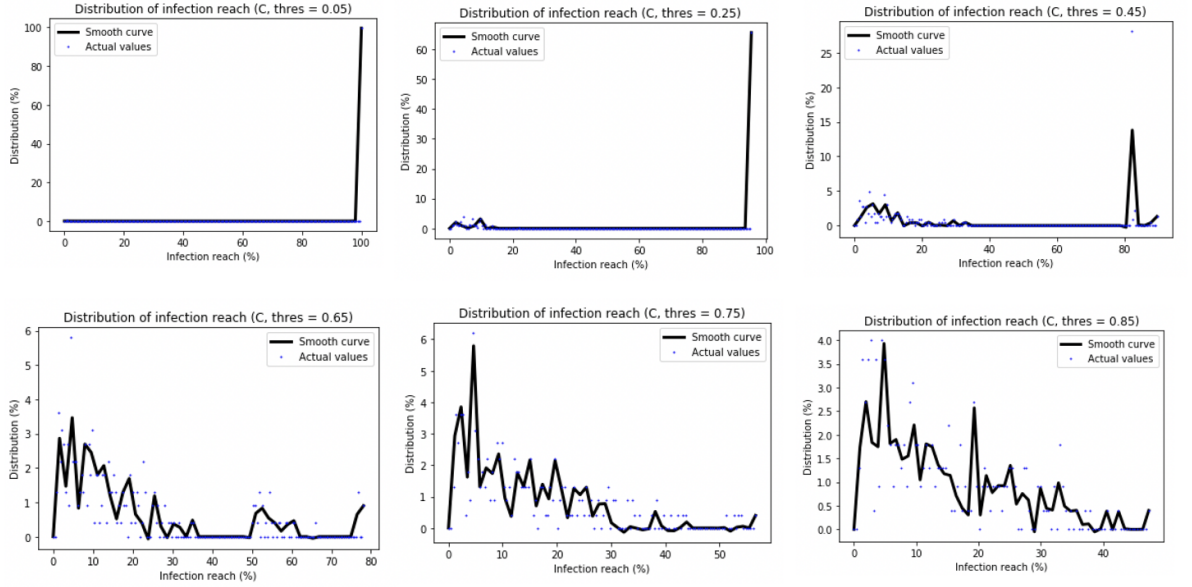
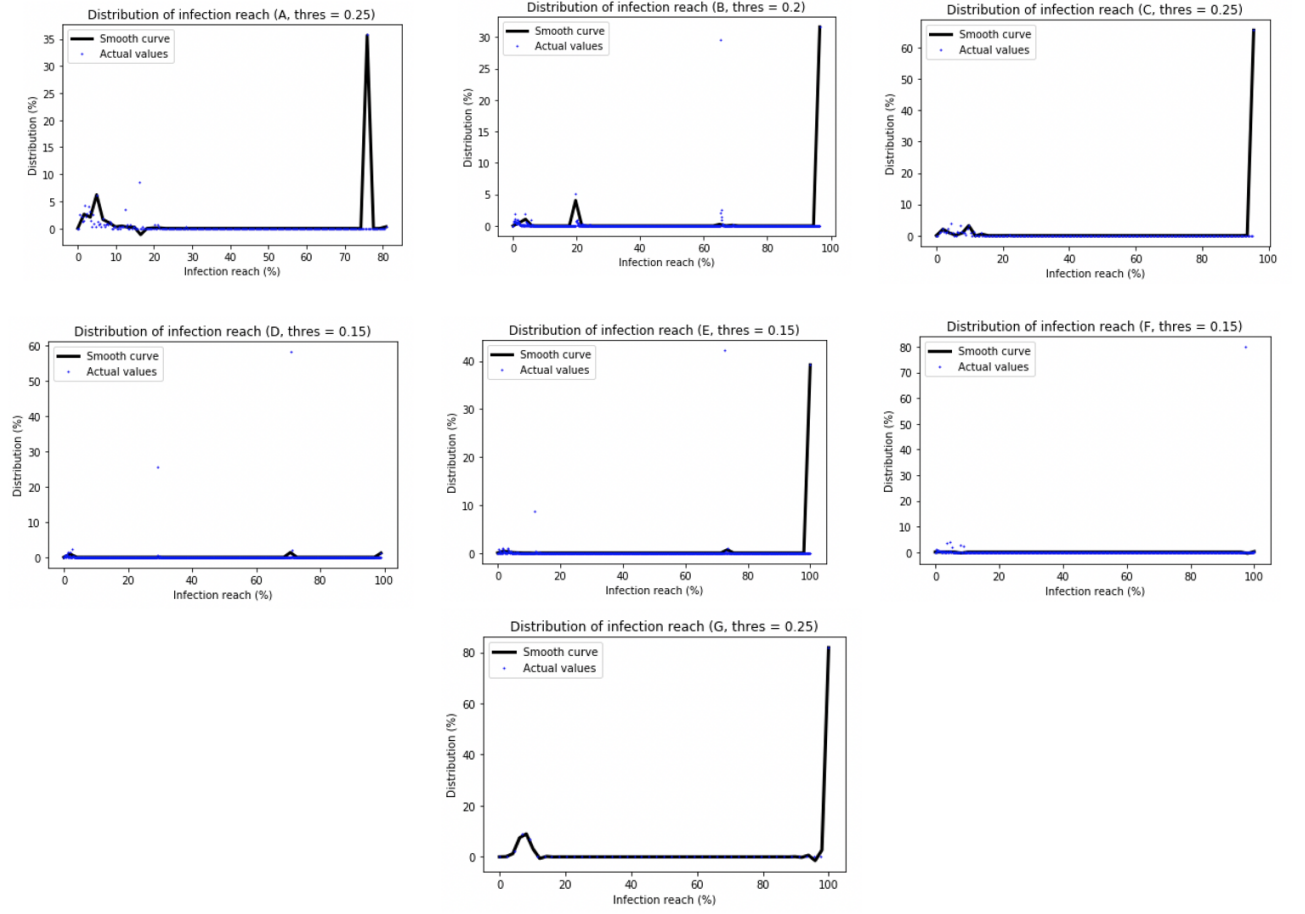


Figure 7: Distribution of infection reach for different relative thresholds  $\alpha$  (Network C): for  $\alpha \leq 0.45$ , we observe the presence of a "critical mass". Any infection reaching this critical number of nodes explodes in cascade and eventually contaminates a high number of nodes. For  $\alpha \geq 0.65$ , the distribution does not exhibit a reach gap.

For a given network of  $n$  nodes, we simulate the deterministic behaviours of contagions by initiating the process at each node successively. We therefore perform  $n$  simulations (See script in Appendix D). Figure 7 displays the distribution of infection reach for network C and different threshold values. For a threshold  $\alpha = 0.05$ , all cascades reach the whole network (100% infection reach for 100% of the simulations). For  $\alpha = 0.25$ , only 65% of the cascades reach the whole network while the others die out before infecting 18% of the agents. For  $\alpha = 0.45$ , the maximal infection reach is 87% but

65% of the contagions do not make it to 40% of the network. Thus, for  $\alpha = 0.25, 0.45$ , the contagion exhibits a clear critical mass (18% for  $\alpha = 0.25$ , 36% for  $\alpha = 0.25$ ). Any infection reaching the critical mass explodes to contaminate a far greater number of nodes. This result provides us with an effective number of nodes to reach so most nodes will be infected. For  $\alpha \geq 0.65$ , the distribution exhibits no clear critical mass but several small gaps. However, the maximal infection reach diminishes as  $\alpha$  rises, which confirms the intuition that the more careful the agents, the less the infection spread. For simple contagions, the critical mass is uninteresting since it is achieved at a single infected node: once the seed node is infected, it infects the whole network (as the graph is connected). For the remaining, we hold  $\alpha$  constant for each network. More particularly, we choose a value of  $\alpha$  that exhibits the critical mass. Figure 8 depicts the distribution of infection reach of each network for the chosen value of  $\alpha$ .



|          | Threshold<br>(alpha) | Critical mass |
|----------|----------------------|---------------|
| <b>A</b> | 0.25                 | 25%           |
| <b>B</b> | 0.20                 | 25%           |
| <b>C</b> | 0.25                 | 18%           |
| <b>D</b> | 0.15                 | 10%           |
| <b>E</b> | 0.15                 | 17%           |
| <b>F</b> | 0.15                 | 15%           |
| <b>G</b> | 0.25                 | 20%           |

Figure 8: Distributions of infection reach for chosen  $\alpha$  values

## 5.2 Critical nodes

We now try to understand how the idiosyncrasies of the seed node (and its local network structure) influence the probability of surpassing the critical mass of infected nodes. Our method is inspired by the recent work *Hurricane Evacuation Behaviour* by Metaxa-Kakavouli et al [11]. They use a statistical method to evaluate the probability

of an individual (living in areas at risk) to evacuate, depending on her/his bonding capital and bridging capitals. We use the simulations performed in Section 5.1. For each network, we fit a single logistic regression to predict a binary variable encoding whether contagion reached the critical mass (1 for yes, 0 for no). As inputs to the models, we used the four social capital metrics introduced in Section 3.2: bonding (first-degree network size, clustering coefficient) and bridging (second-degree network size, span) capitals. We aim at solving:

$$\min_{\theta} \sum_{j=1}^n (p_i - h_{\theta}(\mathbf{x}_i))^2 \quad (3)$$

$$h_{\theta}(\mathbf{x}) = \frac{1}{1 + \exp(-\theta^T \mathbf{x})} \quad (4)$$

where

- $n$  is the network size
- $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$  a set of parameters to optimize.
- $p_i = 1$  if the contagion initiated at node  $i$  surpassed the critical mass,  $p_i = 0$  otherwise.
- $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})$  the social capital vector of node  $i$ :  $x_{i1}$  its first-degree network size,  $x_{i2}$  its local clustering coefficient,  $x_{i3}$  its second degree network size,  $x_{i4}$  its span.

To minimize the error (3), we want  $h_{\theta}(x_i)$  to be close from  $p_i = 1$  (resp.  $p_i = 0$ ) if the contagion initiated at node  $i$  reaches the critical mass (resp. dies out before reaching the critical mass). Moreover,  $h_{\theta}$  is a increasing function of  $x_i$  if  $\theta_i > 0$ . Therefore, if  $\theta_i > 0$ , the higher  $x_i$ , the likelier the infection is to reach the critical mass. That in mind, Figure 9 presents the parameters fit for each network. We scaled and centered the 4 social capital variables so their respective parameters  $\theta_1, \theta_2, \theta_3, \theta_4$  are comparable. Estimates are performed with RStudio v1.1.456 (See script in Appendix C). The first and second degree network size are highly significant and are positively correlated with the surpassing of the critical mass. This suggests that contagion should be initiated with agents that have a high number of friends but also a large variety among friends-of-friends. Thanks to the heterogeneity of their extended circles, they are able to reach different circles. Hopefully, with their amount of friends, they are able to reaffirm information to people within the same circles and,



therefore, allow to trigger the start of behaviour's adoptions within these circles. The influence of the initial clustering coefficient is quite unclear. However, in the networks where it is significant, it is positively correlated with the surpassing of the critical mass. This is in keeping with the idea that the tighter the links within one's friends, the more likely behaviours is to be reinforced and adopted.

| Network  | First degree       | Clustering | Second degree | Span     |
|----------|--------------------|------------|---------------|----------|
| <b>A</b> | 3.3486***          | 1.0663*    | 5.9765***     | -0.37420 |
| <b>B</b> | 5.2852***          | 0.6242***  | 1.4919***     | 0.11400  |
| <b>C</b> | 6.5458***          | -0.18110   | 2.2405**      | -0.66660 |
| <b>D</b> | 3.0801***          | 0.13690    | 1.9788***     | -0.54450 |
| <b>E</b> | 6.2191***          | -0.03702   | 0.68361**     | 0.02008  |
| <b>F</b> | 3.4039***          | -0.15700   | 1.4694***     | -0.3308* |
| <b>G</b> | complete separator |            |               |          |

Figure 9: Fit parameters for the initial local structure properties. Confidence levels: \*\*\*  $p < 0.1\%$ , \*\*  $p < 1\%$ , \*  $p < 5\%$  .

The span of the initial node is never significant except in Network F with a negative coefficient. However, the statistical power of this result could be questioned. Indeed, as shown in Figure 10, we should note that the number of articulation points is low compared with the size of the networks A-F ( $> 220$  nodes). Lastly, for network G, the first degree network size  $x$  is a complete separator (or quasi) of the outcome  $y$  (i.e. there exists  $X$  such that  $x < X \Rightarrow y = 1$  (or  $=0$ ) and  $x > X \Rightarrow y = 0$  (or  $=1$ )).

|                        | A  | B  | C  | D  | E | F  | G |
|------------------------|----|----|----|----|---|----|---|
| Biconnected components | 32 | 23 | 10 | 13 | 5 | 18 | 8 |
| Articulation points    | 24 | 20 | 7  | 11 | 3 | 14 | 6 |

Figure 10: Biconnected components and articulation points

### 5.3 Critical edges

Finally, we investigate where the individuals get their information from. To do so, we look at strength and length of the links responsible for the first infection for a node in the real networks.

For each node (and each simulation), we compute the average strengths of the tie through which the node was infected. As the edge strength distribution is not uniform (see Appendix A), we scale the distribution so all strength values are comparable: if the network has  $x$  edges of strength  $y$  and  $2x$  edges of strength  $2y$ , if the edges of strength  $2y$  are twice more responsible for the node infections than those of strength  $y$ , both edge types are considered to have the same contribution in the node infections. We find that the distribution of the tie strengths through which each individual was infected has a prominent peak next to zero for the networks A-C and E-F, and between 0.25-0.85 for the networks D and G (Figure 11). While the former indicates that in the majority of cases, an individual is contaminated through ties of weak strengths, the latter suggests that individuals can be considerably contaminated through ties of intermediate strength.

Lastly, we carry out a similar investigation for the length of the ties through which the nodes are infected. Unfortunately, the range of edge lengths is very small: the vast majority of the ties have a length between 2 and 4 . The results are then difficult to interpret (See Appendix B).

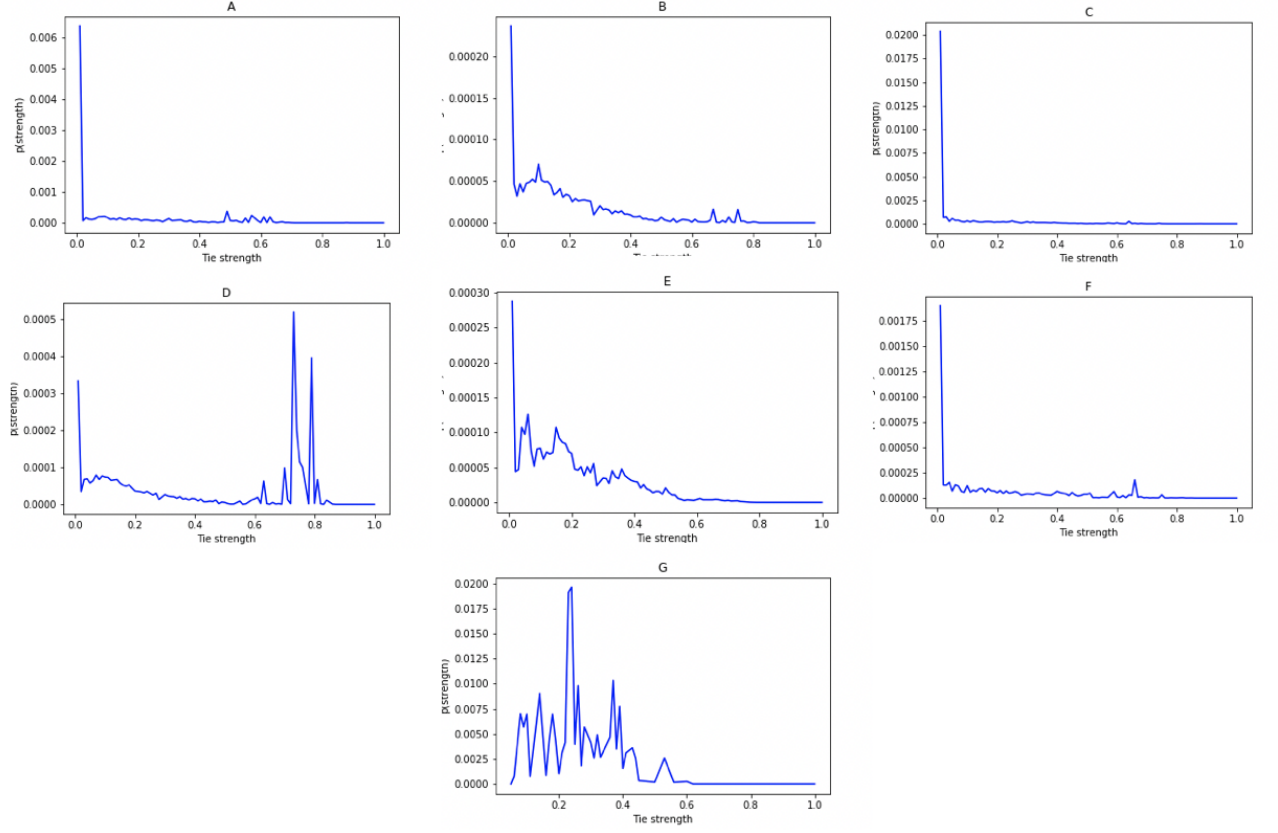


Figure 11: Distribution of strengths of the links responsible for the infection of a node

## 6 Conclusion

We have presented an experimental study that highlights key patterns of complex contagion in 7 Facebook small-world networks. It uses simulations to exhibit the presence of a relatively small critical mass over which the infection is ensured to reach the vast majority of nodes. From this observation, it employs an original statistical method to identify where one should initiate an infection to maximize its chances to attain most of the network. Lastly, it dives into the distribution of strengths (and lengths) of the links responsible for the node infections. Although the 7 networks have different characteristics (size, global clustering coefficient, degree distribution...), we have shown that they all satisfy the small world's criteria. For these small worlds, we have found that the chances of reaching the critical mass are significantly higher when initiating the process at a node with high degree and high second-degree network size. In line with Centola and al. (2007), we have validated the Granovetter's "strength of

weak ties” hypothesis for some networks while showing that it does not apply to all [2].

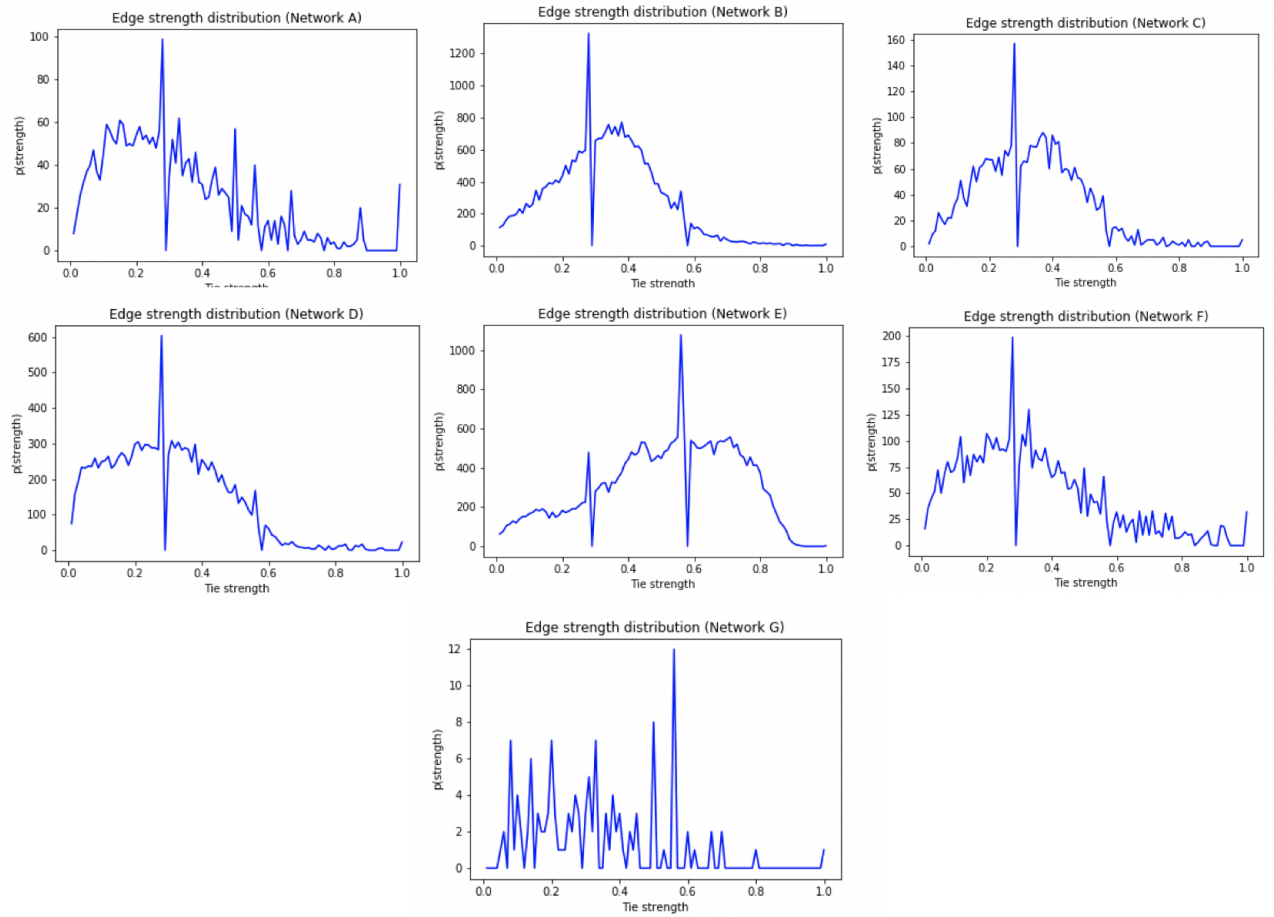
By defining the strength of a tie as the relative neighborhood overlap of its two endpoints, we implicitly assumed that the relationship between two linked agents is embedded with trust. However, two individuals could be tightly bounded (many common friends) but hate each other. In this case, the individuals would be unwilling to adopt each other’s behaviours or trust the information conveyed by each other. The importance of these negative influences in contagion patterns is briefly discussed in [16].

## References

- [1] Mark S. Granovetter (1973), *The Strength of Weak Ties*, American Journal of Sociology 1973 78:6, 1360-1380.
- [2] Damon Centola and Michael Macy (2007), *Complex Contagions and the Weakness of Long Ties*, American Journal of Sociology. Vol. 113, No. 3 (November 2007), pp. 702-734.
- [3] Small-world experiment: [https://en.wikipedia.org/wiki/Small-world\\_experiment](https://en.wikipedia.org/wiki/Small-world_experiment)
- [4] Duncan J. Watts and Steven H. Strogatz (1998), *Collective dynamics of ‘small-world’ networks*, Nature, Vol. 393 (1998), p. 440–442
- [5] Mark D. Humphries and Kevin Gurney (2008), *Network ‘Small-World-Ness’: A Quantitative Method for Determining Canonical Network Equivalence*, PloS one, 3(4), e0002051.
- [6] Renaud Lambiotte (2019), *Networks Notes*, Lecture Notes for C5.4 Networks module at University of Oxford, HT 2018-2019.
- [7] Clay Shirky (2008), *Here Comes Everybody: The Power of Organizing Without Organizations*, [https://uniteyouthdublin.files.wordpress.com/2015/01/here\\_comes\\_everybody\\_power\\_of\\_organizing\\_without\\_organizations.pdf](https://uniteyouthdublin.files.wordpress.com/2015/01/here_comes_everybody_power_of_organizing_without_organizations.pdf)
- [8] Vladimir Barash, Christopher Cameron, Michael Macy (2012), *Critical phenomena in complex contagions*, ScienceDirect, Social Networks, Vol. 34, Issue 4, October 2012, p. 451-461.

- [9] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási (2007), *Structure and tie strengths in mobile communication networks*, PNAS May 1, 2007 104 (18) 7332-7336.
- [10] David Easley and Jon Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Chapter 3 - Strong and weak ties*, <http://www.cs.cornell.edu/home/kleinber/networks-book/networks-book-ch03.pdf> Cambridge University Press (2010)
- [11] Danaë Metaxa-Kakavouli, Paige Maas, Daniel P. Aldrich (2018), *How Social Ties Influence Hurricane Evacuation Behavior*, <https://research.fb.com/publications/how-social-ties-influence-hurricane-evacuation-behavior/> Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)
- [12] Kristen M. Altenburger, Johan Ugander (2017), *Bias and variance in the social structure of gender*, <https://arxiv.org/abs/1705.04774>
- [13] Definition (Articulation point), [https://networkx.github.io/documentation/networkx-1.10/reference/generated/networkx.algorithms.components.biconnected.articulation\\_points.html](https://networkx.github.io/documentation/networkx-1.10/reference/generated/networkx.algorithms.components.biconnected.articulation_points.html)
- [14] James Moody and Douglas R White, *Structural cohesion and embeddedness: A hierarchical concept of social groups*, American Sociological Review (2003), 103–127.
- [15] SNAP - Social Networks: Facebook <http://snap.stanford.edu/data/ego-Facebook.html>
- [16] Vladimir Barash (2011), *The dynamics of social contagion*, <http://www.vlad43210.com/thesis.pdf>

## A Distribution of tie strength



## B Tie length

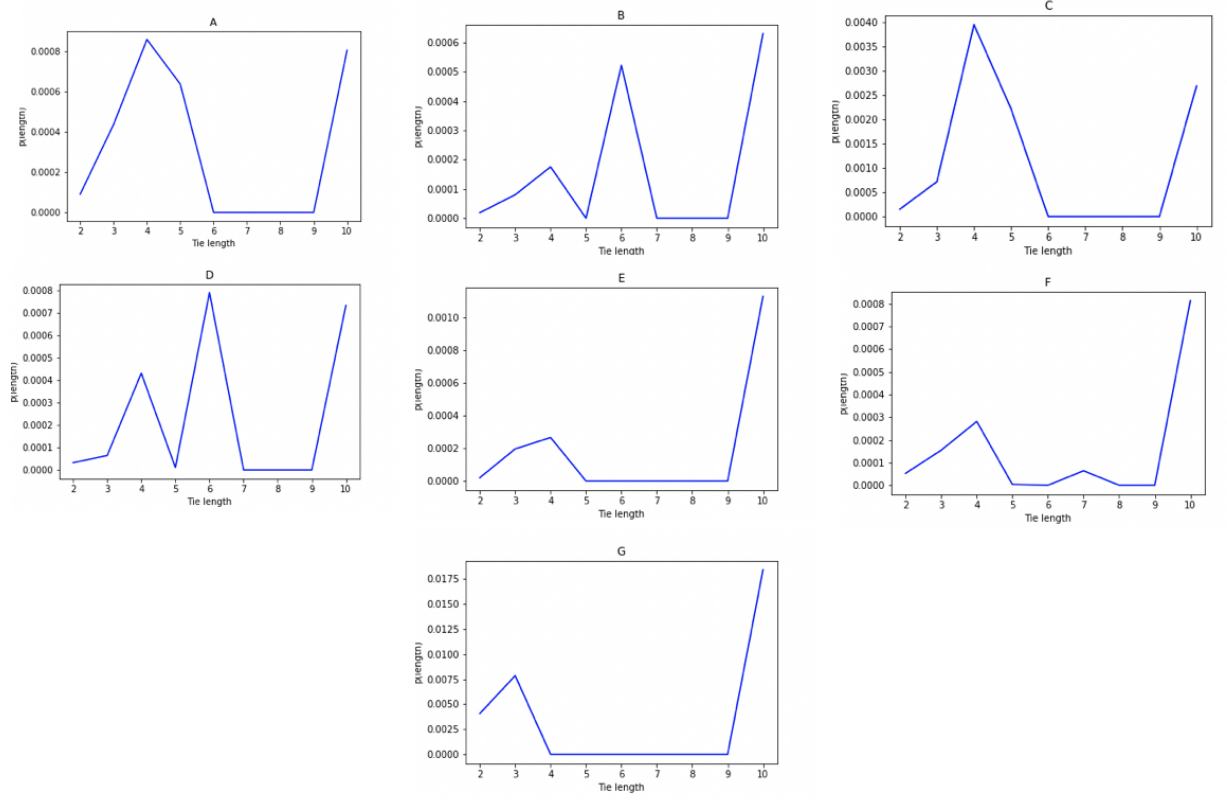


Figure 12: Distribution of lengths of the links responsible for the infection of a node

## C Script (Logistic Regression)

## D Script (Simulations)