Threshold Effects on Diffusion processes

Pablo Bello
Linköping University
pabbe829@student.liu.se

October 2020

Introduction

Diffusion processes constitute a fundamental part of social life. From the spread of social norms and fads to epidemics, diffusion is what makes societies able to be consolidated and open to social change at the same time (Centola 2015). Social networks are the driving structure of diffusion processes and determine them in many ways. To realize this, we do not have to look further than the COVID-19 pandemic. The only effective way that humankind has, so far, to combat the spread of the virus is to destroy part of the social tissue that unites us all. In other cases, such as the uprise of bestsellers, the ways in which social networks come into play are not so obvious, but this does not mean that they are not equally important. Similarly, social conventions, which are the building blocks of society (Searle 2010) would not be in place without a social scaffolding through which to spread.

Models of diffusion abound in the literature, with many different features adapted to the substantive dynamics that they aim to understand. However, all of them include some kind of rule that determines when an individual changes his state, from healthy to infected, non-adopter to adopter or inactive to active. This rule is often a threshold, that is, a number of others to which an individual is connected that is required to trigger adoption in the individual. As Centola puts it, there are fundamentally two types of thresholds:

"Thresholds can be expressed in two ways— as the number (Granovetter 1978) or the fraction (Watts 2002) of neighbors that need to be activated. The conceptual distinction reflects an underlying (and often hidden) assumption about the influence of adopters. Fractional thresholds model contagions in which both adopters and non-adopters exert influence, but in opposite directions. [...] In contrast, numeric thresholds model contagions in which non-adopters are irrelevant." (Centola 2018)

The goal of this paper is to study how numeric and fractional thresholds condition the spread of innovation through a network. For this purpose, I utilize a real network, the friendship network of a Glasgow High School (Pearson and West 2003). The structural properties of this network are modeled in order to simulate networks that resemble the real one. A stochastic actor-oriented model (T. A. Snijders, Bunt, and C. E. Steglich 2010) is used as a way to generate what can be considered as an empirically informed agent-based model (T. A. B. Snijders and C. E. G. Steglich 2015).

Theory

Examples of the use of numeric as well as fractional thresholds abound in the literature but, arguably, the most influential ones in sociology are Schelling's (2006) and Granovetter's (1978) threshold models. In his work, Granovetter explores a number of different diffusion processes for which one of the threshold rules applies. For instance, he assumes that when it comes to joining a riot, adopting and innovation or making educational decisions what matters is the proportion of one's reference group that has adopted the behavior in question. For the spread of rumors and diseases, leave a social occasion or migrate to a new country, what people really care about is the absolute number and not the proportion of his reference group that has done it before them. On a footnote, Granovetter admits taking the idea of behavioral thresholds for contagion from Schelling's models of segregation. Schelling also leaves unquestioned the issue of fractional versus numeric thresholds, using one or the other arbitrarily. After these, many other threshold models followed (see for instance Bicchieri 2017; Chwe 1999; Kuran 1997; Mackie 1996; M. W. Macy 1991).

In general, the implications of choosing between numeric and fractional thresholds were not examined. The decision was based on intuitions about the individual-level processes of decision making that drive the spread of a particular social phenomenon. It was not until computational methods allowed the replication of emergent processes that start thinking about how to define thresholds gained more relevance. Now, the consequences of this decision could be actually observed as the diffusion unfolds. Nonetheless, this dimension of thresholds is still often neglected, even in models similar to the one presented here (Greenan 2015; Iacopini et al. 2019; Kempe, Kleinberg, and Tardos 2015).

To my knowledge, the only computational exploration of the consequences of numeric versus fractional thresholds has been done in Centola's work (Centola 2018; Centola and M. Macy 2007). In his simulations, Centola first generates a continuum of networks that go from a clustered scale-free network to a randomly wired scale-free network. The process starts by setting a group of initial adopters and then observing whether the diffusion cascades or not. When the threshold is fractional, behavior only spreads in highly clustered networks because otherwise, the hubs do not get infected. If instead, the threshold is numeric, the success of diffusion is not contingent upon the level of clustering in the network. In short, what happens is that hubs, which are highly connected nodes, are very susceptible to get infected when thresholds are numeric simply because they are in contact with many other nodes.

One issue in Centola's work is that, although scale-free networks are common in many realms (Barabási and Albert 1999) they are not especially representative of social networks (Watts 2004). Social networks, typically have degree distributions with less variance than scale-free networks. For instance, if we think of friendship networks, we can admit that some individuals have more friends than others, but having more than, let us say, 50 friends, would be implausible (with a strict definition of what friendship is). Thus, the difference between the number of friends of the least popular individual in a network and the most popular one will never be as sizeable as differences in the number of links referencing web pages on the internet. The reason why Centola uses scale-free networks is that the more variance there is in the degree distribution the bigger the difference between fractional and numeric thresholds. Here, I use a real network as a template for simulating the spread of innovation so results can be more easily generalized to real-world social networks. Another difference with Centola's approach is that instead of varying the clustering in the network

I focus instead on the effects that different threshold levels and speeds of change have on the diffusion process.

Method

Stochastic Actor Oriented Models (SAOMs) are one of the most consolidated approaches to the study of dynamic social networks. These models permit the statistical analysis of the coevolution network and behavior, thus making possible to disentangle, for instance, the effects of influence and homophily (C. Steglich, T. A. B. Snijders, and Pearson 2010). Another, less explored, use for SAOMs is the simulation of networks with the goal of exploring micro-macro links:

"The actor-based nature of the model implies that changes in the network are modeled as choices by the actors. This leads to a model combining agency and structure, which is well suited for expressing theories based on purposeful behavior by social actors conditioned by their network context, but also for exploring the macro-level consequences of these theories" (T. A. B. Snijders and C. E. G. Steglich 2015).

SAOM has been successfully used as a simulation tool in studying the effects of peer influence and homophily on adolescent smoking (Lakon et al. 2015; Schaefer, j. a. j., and Haas 2013) as well as the diffusion of beliefs (C. Steglich n.d.) and innovation (Greenan 2015).

In this paper, I explore how the individual rules that actors follow to adopt a new behavior affect the spread of such behavior through the network. The spread of innovation in a network can be regarded as an aggregated micro feature (T. A. B. Snijders and C. E. G. Steglich 2015) since it is a direct consequence of the actions of the actors in the network. Although it is not a strictly emergent macro feature, the difficulty of deriving the aggregated consequences of the type of threshold in an analytical manner justifies its study through simulations.

Nonetheless, some properties of the numeric and fractional thresholds can still be analytically derived for the individual level. Let us define a numeric threshold (N), that can range from 1 (a simple contagion) to infinite, and a fractional threshold (F), that can range from 0 to 1. In a network with a constant degree (d), if N/F = d then the numeric and fractional definitions of the threshold are equivalent. In a network with varying degrees, for nodes with d = N/F this still holds. However, for nodes with d ¡ N/F the probability of adoption is higher with a fractional than with a numeric threshold. Nodes with d ¿ N/F are more susceptible to adopt under a numeric than a fractional threshold. On an aggregated level, the distribution of degrees around the quantity N/F will determine the outcome of the diffusion process in ways that are impossible to trace since this relationship is also dependent on the structure of the network.

It does not exist a unique way to define N/F so my approach here is to estimate it as the results of a random diffusion process that does not take into account the network structure. I start by assigning an equal probability to each node of adopting at time 1 (p = 0.2). Apart from convenient, experimental research has shown that an initial mass of around 20% of the population is required to trigger certain social diffusion processes (Centola 2018; Centola and Baronchelli 2015). In time 2, each node that was not an adopter in time 1 has an equal probability of having adopted the behavior by then (p = 0.4). From this random process, in which the structure of the network plays no role, I estimate, via SAOM an N/F 7.6 (the inverse of totExposure/avExposure, which are the RSiena effect for N and F respectively). This might not be the only method to estimate N/F, and that is likely to have a significant impact on the results of the simulations but it is an improvement compared to previous studies (Centola 2018).

In order to arrive at the results presented here, I go through three main steps: estimation, simulation, and replication. Estimation. I estimate two SAOMs until reaching convergence, one including a fractional and other with a numeric threshold for adoption. The RSiena total exposure effect compares to the numeric definition of threshold while the average exposure effect compares to the fractional specification. Each model also includes effects for transitivity and reciprocity This way, the simulated networks will resemble (at least on these two key characteristics) the observed network(goodness of fit test are also run to make sure that the simulated networks are similar to the observed network).

Simulation: For the simulations, I keep the network constant. Thus, actors do not have any opportunities for changing their ties while the adoption process unfolds. I manipulate the total and average exposure effects in order to generate different levels of adoption at the end of the process. This can be thought of as increasing or decreasing the individual thresholds for adoption. The bigger the effect the lower the threshold and vice versa. For each manipulation, I generate a hundred networks and observe the number of actors that adopt the behavior at the end of the process.

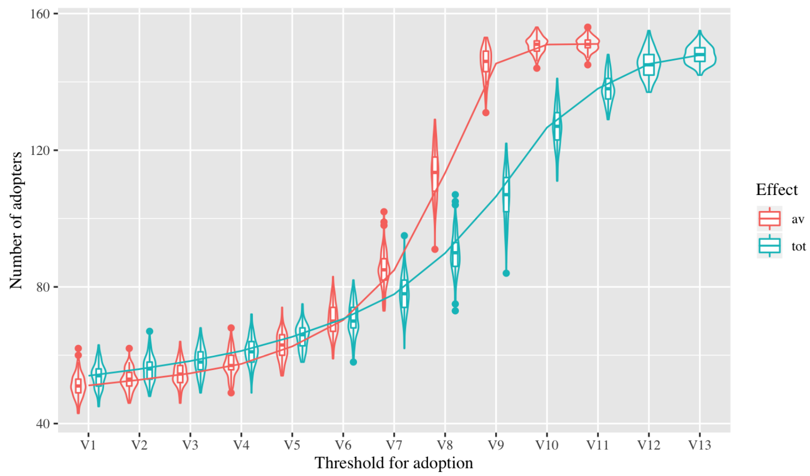
Replication: To make sure that the results are not an artifact of the initial position of the adopters I permute them a hundred times and run the simulations again, keeping everything else constant.

Results

Figure 1 presents the results for the simulations with each manipulation of the threshold. The red line corresponds to the average exposure model (fractional threshold) and the turquoise one to the total exposure model (numeric threshold). The point at which both models generate the same number of adoptions (V6) is the initial estimate. If we make the effects smaller (but always keeping N/F constant) then adoption decays slightly more so in the average exposure model than in the total exposure model. Similarly, if we increase the effects (lowering the threshold for adoption) the number of adopters increases. The adoption curve is s-shaped for both models. More importantly, the average exposure model presents a steeper curve of adoption than the total exposure model. By V11 the network is saturated. Not all the nodes have adopted the new behavior because there are isolates but almost anybody else in the network is an adopter. For the total exposure model, it takes a lower threshold (V13) to get to the flat upper part of the S-curve and even then, levels of adoption are not as high as for the average exposure model. These results are robust to permutations in the initial position of the nodes (see figure 4 in the appendix).

Instead of manipulating the threshold values, we can leave them constant while increasing or decreasing the opportunities for adoption (the adoption rate function in the models). Figure 2 represents this process. Now, the shape of the adoption curve resembles the increasing section of a negative quadratic function. The difference between models still holds, being the average exposure model the one that generates more and faster adoptions. Thus, the superior efficacy of an

Figure 1. Simulated diffusion process on the Glasgow network



average threshold for spreading innovation, under this particular N/F condition, does not depend on the threshold size or the number of opportunities for change.

Figures 3 and 4 show representative networks for each manipulation. I consider representative networks those whose final number of adopters equals the mean number of adopters for the one hundred simulated networks of each manipulation. On the first network (V1) which has the highest threshold, there are only a few activated nodes (in red), which are mainly initial adopters. The V11 network for the average exposure model is almost saturated. Only some isolates and peripheral nodes remain inactive (green nodes). For the total exposure model, it takes lower thresholds to get to almost complete saturation of the network.

As explained above, differences in adoption between the numeric and fractional models should be in part stem from the degree distribution of the network. Under equal conditions, poorly connected actors will be more susceptible to adopt with a fractional threshold whilst highly connected actors will become active more easily under a numeric threshold. To see if this is what is happening in the simulations figures 5 and 6 (in the appendix) represent the distribution of the total degree of the nodes by adoption status for some manipulations. When adoption is random (V6) the degree distributions of adopters and non-adopters are pretty similar. If we lower the thresholds (V7 to V13) the two distributions drift apart. The lower the threshold the more non-adopters concentrate on low

Figure 2. Simulated diffusion process on the Glasgow network

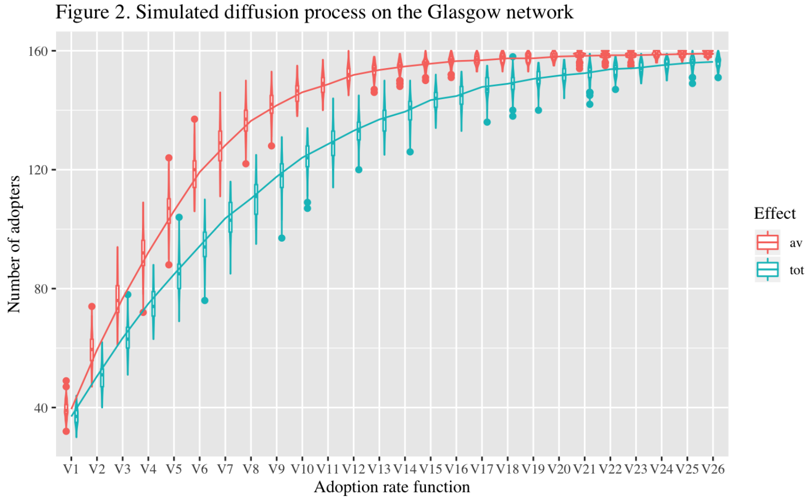
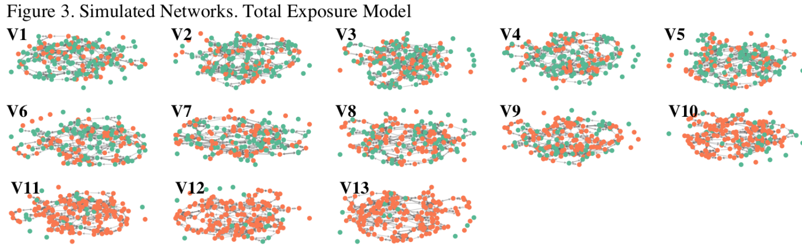
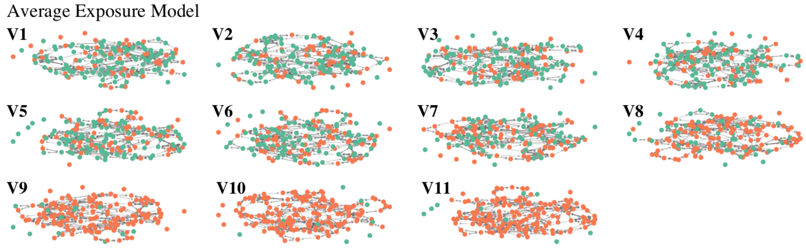


Figure 3. Simulated Networks. Total Exposure Model



Average Exposure Model



degree values (without taking into account those with degree 0 which obviously cannot be influenced). If we now compare the two models, we can observe that, for equivalent levels of adoption, the distribution of non-adopters is more evenly distributed in the average exposure model than in the total exposure model. Although in both model less connected actors are more reticent to adoption, this is accentuated when thresholds are numeric.

Discussion

The key takeaway of this study should be that choosing between a numeric or a fractional definition of a threshold in adoption processes is not trivial. A numeric threshold resembles an individual decision rule in which friends that have not adopted a behavior are not taken into account. Fractional thresholds, on the other hand, can be useful when modeling diffusion processes for which peer-influence can be exerted in opposite directions (to maintain the previous behavior or adopt a new one). Although common sense could induce us to believe that numeric thresholds will always spread behaviors further into the network because non-adopters do not exert any influence, our model shows that this is not always the case. It will depend on the relationship between the value of the fractional and numeric threshold and how the nodes' degree is distributed around it. Exploring the conditions under which numeric or fractional thresholds are more effective in spreading behavior deserves further consideration.

One last conclusion is that even when degree heterogeneity is minimal the selection of threshold type has substantive consequences at the macro level. Previous studies had tested similar models of fractional and numeric threshold effects on diffusion processes in scale-free networks, which have high degree heterogeneity by definition. However, this network structure is not representative of real social networks like the high school network used in this study. Here, we prove that even small amounts of degree heterogeneity, make the threshold definition consequential.

References

Barabási, Albert-László and Réka Albert (Oct. 1999). "Emergence of Scaling in Random Networks". en. In: *Science* 286.5439. Publisher: American Association for the Advancement of Science Section: Report, pp. 509–512. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.286.5439.509. URL: https://science.sciencemag.org/content/286/5439/509 (visited on 04/02/2020).

Bicchieri, Cristina (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. en. Google-Books-ID: v214DQAAQBAJ. Oxford University Press. ISBN: 978-0-19-062205-3.

Centola, Damon (Mar. 2015). "The Social Origins of Networks and Diffusion". en. In: *American Journal of Sociology* 120.5, pp. 1295–1338. ISSN: 0002-9602, 1537-5390. DOI: 10.1086/681275. URL: https://www.journals.uchicago.edu/doi/10.1086/681275 (visited on 04/02/2020).

— (2018). *How behavior spreads: the science of complex contagions*. en. Princeton Analytical Sociology Series. OCLC: on1035756208. Princeton; Oxford: Princeton University Press. ISBN: 978-0-691-17531-7.

Centola, Damon and Andrea Baronchelli (Feb. 2015). "The spontaneous emergence of conventions: An experimental study of cultural evolution". en. In: *Proceedings of the National Academy of Sciences* 112.7, pp. 1989–1994. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1418838112. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1418838112 (visited on 04/03/2020).

Centola, Damon and Michael Macy (Nov. 2007). "Complex Contagions and the Weakness of Long Ties". en. In: *American Journal of Sociology* 113.3, pp. 702–734. ISSN: 0002-9602, 1537-5390. DOI: 10.1086/521848. URL: http://www.journals.uchicago.edu/doi/10.1086/521848 (visited on 04/02/2020).

Chwe, Michael Suk-Young (July 1999). "Structure and Strategy in Collective Action". en. In: *American Journal of Sociology* 105.1, pp. 128–156. ISSN: 0002-9602, 1537-5390. DOI: 10.1086/210269. URL: https://www.journals.uchicago.edu/doi/10.1086/210269 (visited on 04/02/2020).

Granovetter, Mark (May 1978). "Threshold Models of Collective Behavior". en. In: *American Journal of Sociology* 83.6, pp. 1420–1443. ISSN: 0002-9602, 1537-5390. DOI: 10.1086/226707. URL: https://www.journals.uchicago.edu/doi/10.1086/226707 (visited on 03/28/2020).

Greenan, Charlotte C. (Jan. 2015). "Diffusion of innovations in dynamic networks". en. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178.1, pp. 147–166. ISSN: 09641998. DOI: 10.1111/rssa.12054. URL: http://doi.wiley.com/10.1111/rssa.12054 (visited on 04/02/2020).

Iacopini, Iacopo et al. (Dec. 2019). "Simplicial models of social contagion". en. In: *Nature Communications* 10.1, p. 2485. ISSN: 2041-1723. DOI: 10.1038/s41467-019-10431-6. URL: http://www.nature.com/articles/s41467-019-10431-6 (visited on 03/28/2020).

Kempe, David, Jon Kleinberg, and Eva Tardos (2015). "Maximizing the Spread of Influence through a Social Network". en. In: *Theory of Computing* 11.1, pp. 105–147. ISSN: 1557-2862. DOI: 10.4086/toc.2015.v011a004. URL: http://www.theoryofcomputing.org/articles/v011a004 (visited on 04/02/2020).

Kuran, Timur (1997). *Private Truths, Public Lies*. en. Google-Books-ID: HlK-BaiCpSxYC. Harvard University Press. ISBN: 978-0-674-70758-0.

Lakon, Cynthia M. et al. (Dec. 2015). "Simulating Dynamic Network Models and Adolescent Smoking: The Impact of Varying Peer Influence and Peer Selection". en. In: *American Journal of Public Health* 105.12, pp. 2438–2448. ISSN: 0090-0036, 1541-0048. DOI: 10.2105/AJPH.2015.302789. URL: http://ajph.aphapublications.org/doi/10.2105/AJPH.2015.302789 (visited on 03/28/2020).

Mackie, Gerry (1996). "Ending Footbinding and Infibulation: A Convention Account". In: *American Sociological Review* 61.6, pp. 999–1017. ISSN: 0003-1224. DOI: 10.2307/2096305. URL: https://www.jstor.org/stable/2096305 (visited on 01/08/2020).

Macy, Michael W. (Dec. 1991). "Chains of Cooperation: Threshold Effects in Collective Action". en. In: *American Sociological Review* 56.6, p. 730. ISSN: 00031224. DOI: 10.2307/2096252. URL: http://www.jstor.org/stable/2096252?origin=crossref (visited on 04/02/2020).

Pearson, Michael and Patrick West (2003). "Drifting smoke rings: Social network analysis and Markov processes in a longitudinal study of friendship groups and risk-taking, Connections". In: *CONNECTIONS* 25.2, pp. 59–76.

Schaefer, David R., jimi adams jimi, and Steven A. Haas (Oct. 2013). "Social Networks and Smoking: Exploring the Effects of Peer Influence and Smoker Popularity Through Simulations". en. In: *Health Education & Behavior* 40.1_suppl, 24S–32S. ISSN: 1090-1981, 1552-6127. DOI: 10.1177/

1090198113493091. URL: http://journals.sagepub.com/doi/10.1177/1090198113493091 (visited on 04/02/2020).

Schelling, Thomas C. (Oct. 2006). *Micromotives and Macrobehavior*. en. Google-Books-ID: DenWKRgqzWMC. W. W. Norton & Company. ISBN: 978-0-393-06977-8.

Searle, John (Jan. 2010). *Making the Social World: The Structure of Human Civilization*. en. Google-Books-ID: kz6R0eDZ5OEC. Oxford University Press. ISBN: 978-0-19-974586-9.

Snijders, Tom A. B. and Christian E. G. Steglich (May 2015). "Representing Micro–Macro Linkages by Actor-based Dynamic Network Models". In: *Sociological Methods & Research* 44.2. Publisher: SAGE Publications Inc, pp. 222–271. ISSN: 0049-1241. DOI: 10.1177/0049124113494573. URL: https://doi.org/10.1177/0049124113494573 (visited on 04/02/2020).

Snijders, Tom A.B., Gerhard G. van de Bunt, and Christian E.G. Steglich (Jan. 2010). "Introduction to stochastic actor-based models for network dynamics". en. In: *Social Networks* 32.1, pp. 44–60. ISSN: 03788733. DOI: 10.1016/j.socnet.2009.02.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378873309000069 (visited on 04/02/2020).

Steglich, Christian (n.d.). "Why echo chambers form and network interventions fail: Selection outpaces influence in dynamic networks1." en. In: (), p. 31.

Steglich, Christian, Tom A. B. Snijders, and Michael Pearson (Aug. 2010). "8. Dynamic Networks and Behavior: Separating Selection from Influence". In: *Sociological Methodology* 40.1. Publisher: SAGE Publications Inc, pp. 329–393. ISSN: 0081-1750. DOI: 10.1111/j.1467-9531.2010.01225.x. URL: https://doi.org/10.1111/j.1467-9531.2010.01225.x (visited on 04/03/2020).

Watts, Duncan J. (Feb. 2004). *Six Degrees: The Science of a Connected Age*. en. Google-Books-ID: 1gueFWR7qjoC. W. W. Norton & Company. ISBN: 978-0-393-07612-7.

Appendix

Figure 4. Simulated diffusion process on the Glasgow network (Replications)

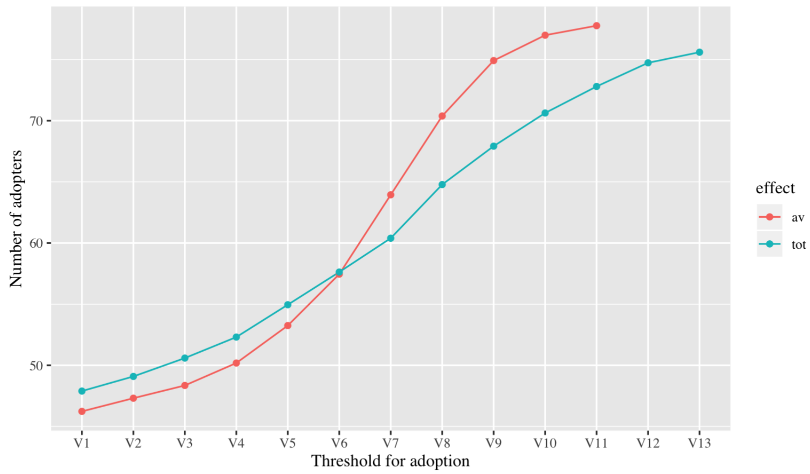


Figure 5. Density plots for the total exposure model

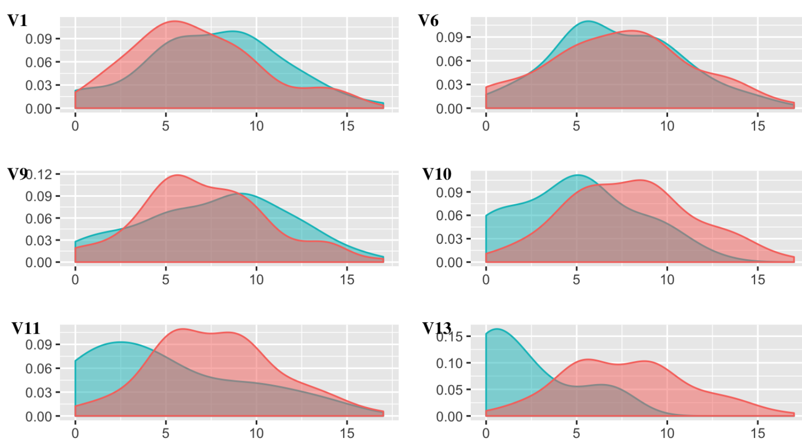


Figure 6. Density plots for the average exposure model