IMPACT OF HOMOPHILY ON DIFFUSION DYNAMICS OVER SOCIAL NETWORKS

Mustafa Yavaş
Gönenç Yücel
Socio-Economic System Dynamics Research Group
Bogazici University Industrial Engineering Department
34342 Bebek Istanbul Turkey
mustafa.yavas@boun.edu.tr, gonenc.yucel@boun.edu.tr

KEYWORDS

Homophily, diffusion, threshold models, social networks, formal models, simulation, segregation.

ABSTRACT

The purpose of this study is to find out under what conditions homophily reinforce the diffusions over social networks or undermines them. To realize this aim, formal modeling approach is utilized and an Agent-Based Model is constructed. Afterwards, diffusion of a non-sticky innovation is investigated with the experiments having varying homophily levels in a social network with two distinct kinds of agents as the primary control variable. The results show that (i) homophily reinforces itself (ii) looking at the macro-behavior of the diffusion, initial increases in the level of homophily has a positive effect on adopted fraction of the population whereas further increases have a negative impact, and (iii) looking at the micro-behavior of the diffusion, increasing homophily can result in local maxima even the macro trend is decreasing. Connectedness and average degrees interacting with social persuasion are the two explanatory remarks in the course of investigating the impact of homophily. As a by-product, the model is also capable of capturing the segregation dynamics over social networks. Future research involves allowing the adopted innovation to lead to value homophily, exploration of the different diffusion initiation types and adoption heuristics.

INTRODUCTION

People tend to form social links more with people who are similar to them and this refers to the principle called *homophily*: "Birds of a feather flock together" (McPherson, Smith-Lovin, and Cook 2001). Homophily is ubiquitous, it can be observed in almost all kind of social network ties such as marriage, friendship, membership in an organization, information exchange, trade and business transactions, and the like. People consider the other party's race, gender, age, religion, culture, education level, and many other personal properties when it comes to social interactions. Being that much pervasive, it is one of the key explanations why some characteristics are localized in social space. In addition to having an inclination to bond more, people meet more frequently and more intensely with friends alike. Furthermore, close friends, who are usually friends alike in some certain qualities, have more influence on us than dissimilar ones (McPherson et al. 2001). Taking all into account, localization of a social aspect is realized via homophily principal in every kind of social networks.

Homophily determines the network of a person. Hence, along with localization, it is also significant in access to information (Choudhury, Sundaram, and John 2010), spread of behaviors, innovations, ideas, and even states of health such as obesity (Christakis and Fowler 2007), opinion and norm formation (Centola, Willer, and Macy 2005), intergroup inequalities (Paul DiMaggio and Filiz Garip 2012; F Garip and P DiMaggio 2011), and so forth. Among these, the interplay between homophily and diffusion covers the most of the recent debate (Jackson and Lopez-Pintado 2011). In this paper, we will only focus on the effect of homophily on diffusion dynamics.

Homophily has a significant impact on the diffusion patterns over a social network via two distinct processes. First, homophily has an impact on the development of a network, hence on its topology. Second, people are more effective in exerting social influence on people alike. However, it is unclear whether homophily is in favor of diffusion or works against the spread. On the one hand, it can promote the diffusion via attaining the critical mass faster than otherwise (Centola 2010; Rogers 2003). On the other hand, homophily can lead the innovation, the norm or the behavior that is supposed to diffuse to become localized in some clusters within the network (Centola 2011; Rogers 2003). Hence, it is not straightforward to see the overall impact of homophily on the diffusion patterns.

The ultimate purpose of this study is to find out under what conditions homophily promotes or prevents the diffusion of an innovation which can be an idea, an attitude, a behavior, a product, and so on.. To do so, a dynamic simulation model will be used to address how the diffusion patterns change when homophily level in a

Proceedings 27th European Conference on Modelling and Simulation ©ECMS Webjørn Rekdalsbakken, Robin T. Bye, Houxiang Zhang (Editors)
ISBN: 978-0-9564944-6-7 / ISBN: 978-0-9564944-7-4 (CD)

social network varies. As a by-product, we will also have the chance to show how homophily changes the network topology.

This article is structured as follows; the following section will elaborate on the problem. In Section 3, the simulation model will be explained. Afterwards, Section 4 presents the outcomes of the simulation experiments conducted to reveal the impact of homophily. In Section 5, interesting results and the factors that give rise to them are discussed. Lastly, Section 6 is devoted to key findings of this study and its limitations, and future research directions.

PROBLEM DEFINITION

Diffusion dynamics is a hot topic in various academic research environments and the problems are ubiquitous. Depending on the nature of the innovation as well as the context within which it is expected to diffuse, the diffusion problems can come in very different forms. However, there are three significant aspects that define diffusion problems in which homophily has a strong potential to influence dynamics. These three aspects, which will be discussed below, define the type of diffusion problems considered in the scope of this study.

Firstly, we will consider an innovation that is quite social in the sense that an individual's adoption is very sensitive to the adoption behavior of the others around. In other words, we will only consider cases of diffusion where social influence, rather than pure personal preference, is the key determinant of a person's adoption decision.

Secondly, it is assumed that the influence of the people alike is significantly stronger than other social contacts in adoption decisions of individuals. Simply put, homophily is active in adoption-related social influence. As we talk about "people alike", it is important to mention a relevant distinction in similarity and related types of homophily: *status homophily* is based on similarities in ascribed status such as race, ethnicity, age or gender and acquired status religion, education, occupation etc., whereas *value homophily* is based on values, innovations, and beliefs (Lazarsfeld and Merton 1954). Status homophily seems like one of the most effective factors leading to strongest divides in sociodemographic space (McPherson et al. 2001). With that motivation, we will focus on status homophily in this study.

Thirdly, the nature of the innovation is also an important aspect. In these types of studies, it is important to distinguish two distinct kinds of innovations: Innovations that can accumulate inertia can be called as sticky innovations, which once adopted, cannot be changed for a long time. On the other hand, innovations with negligible or no inertia can be called as non-sticky innovations, which can be changed frequently. Studying sticky innovations as the diffusing phenomenon decrease the explanatory power of a model that aims to study the impact of homophily due to the fact that inertia has as a confounding effect. Instead, non-sticky innovations would maximize the potential to investigate the impact of homophily. Thus, we will focus on the dynamics of the diffusion of a non-sticky innovation.

To summarize, we will focus only on status homophily and its impact on the diffusion of a non-sticky innovation. For the sake of simplicity and not lose explanatory power, the social network that we will focus on will have two kinds of people with respect to a hypothetical status which leads them to form homophilious ties: green and red. We will assume that this status is a permanent one such as race, or gender. We will further assume that this status is essential to adopting the innovation that is supposed to diffuse. Only the individuals of the same status will have an influence on each others' adoption and quitting decisions.

Taking such a problem into account, it can be stated that there are no global but local interactions and the micro rules of them are well-known; individuals form ties more with people alike and they are influenced only by them. Furthermore, each individual has heterogeneous attributes that are closely related to diffusion process such as having adopted the innovation or not, qualities relating the network such as number of friends, her exact position in the topology, fraction of similar friends around her, etc. Additionally, due to aforementioned conflicting effects of homophily, it is not straightforward to deduce the macro pattern in a diffusion study with homophily being included both in evolution of network and affecting the adoption process and a computational model would be helpful. Taking all those into account, Agent-Based Modeling (ABM) is evaluated as a promising approach for our study (Epstein and Axtell 1996; Gilbert and Troitzsch 2005; Railsback and Grimm 2011). An agent-based model that depicts the diffusion environment that is defined by our main assumptions will be developed and it will be used to conduct experiments to analyze the overall effect of homophily on diffusion dynamics.

MODEL DESCRIPTION

To address the impact of homophily, two distinct phases should be completed. First, a network which has already some measurable level of homophily is needed to be built. After such a given network, initiation of the diffusion by choosing early adopters should take place and we will let the diffusion begin. In this regard, the model will be described in two parts, i.e. network evolution and diffusion.

Homophilious Network Evolution

The evolution process of the homophilious network is inspired by the famous "Segregation Model" (Schelling 1971). The process starts with a *small world* network (Wilensky 2005). On this network, each agent has a

threshold value for her "happiness level", as in Schelling's model, related to the fraction of the similar friends in the agent's network neighborhood. "Unhappy" agents will make ties with agents alike, and break ties with others.

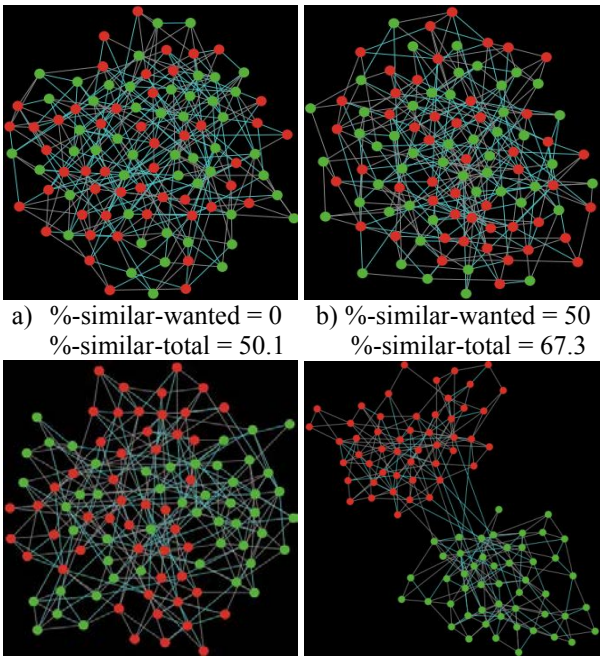
%-similar-wanted is the model parameter that represents this personal threshold. After setting a value for *%-similar-wanted*, the following rules will be applied:

> *<u>Homophilious network evolution routine:</u>*
>
> *Create a "small world" network*
>
> *WHILE average "happiness level" is less than %-similar-wanted*
>
> *{Ask each agent:*
>
> *IF her "happiness level" is less than %-similar-wanted*
>
> *THEN form a new tie with a similar person from your friends' friendship networks*
>
> *AND dissolve a tie with a dissimilar friend of yours*
>
> *ELSE do nothing*
>
> *Calculate average "happiness level"}*
>
> *IF the network is connected THEN do nothing ELSE return to the beginning and create a new network*

In the above routine, "happiness level" refers to the fraction of similar friends in an agent's friendship network. Note that all agents have the same threshold value for "happiness level" and that is equal to *%-similar-wanted*. The above heuristic ensures that the average degree of the social network stays the same. Nevertheless, in the end, agents are still heterogeneous in terms of "happiness level", average degree (i.e. number of friends), network statistics values of each agent's friendship network such as clustering coefficient, and centrality values such as betweenness-centrality, closeness-centrality, and so forth.

As the network evolves according to the set *%-similar-wanted* value, the homophily level of the network is measured with the average of fractions of similar friends in agents' friendship networks (*%-similar-total*). Although, to the best of our knowledge, there is no well-known and widely accepted measure for homophily, there are examples of the one that is used here (Jackson 2008). Note that in the beginning, due to random assignments of agents, on average, *%-similar-total* is 50% and it changes as the network evolves homophiliously with respect to value of *%-similar-wanted*.

In *Figure 1*, we see a similar result as in Schelling's Segregation Model: as *%-similar-wanted* increases, network tends to become more segregated.



a) %-similar-wanted = 0 %-similar-total = 50.1 b) %-similar-wanted = 50 %-similar-total = 67.3

c) %-similar-wanted = 75 %-similar-total = 86.9 d) %-similar-wanted = 85 %-similar-total = 97.7

Figure 1: Instances of social networks with 50 nodes per status and average degree of 6.

Diffusion

Besides its role in network evolution, homophily is also influential during the diffusion. However, the network structure will stay the same since homophily will be present only in the adoption decisions. In other words, we assume that the adopted behavior does not create homophilious tie formation or dissolution.

Diffusion starts with the initiation of the early adopters and takes off as people make decisions whether to adopt the innovation or not via social influence. The adoption mechanism in this study is an example of *threshold models* (Granovetter 1978). The most important ingredient of adoption decision is "adoption threshold". As its name implies, it is the threshold fraction of similar friends who already adopted the innovation. After the initiation step, in each round, the following heuristic will begin to run until the adoption dynamics reach equilibrium:

> *<u>Adoption routine:</u>*
>
> *Ask each agent:*
>
> *IF (#-of-similar-friends-adopted / #-of-similar-friends) is greater than adoption threshold THEN*
>
> *IF he has not adopted THEN make her adopt it ELSE do nothing*
>
> *ELSE*
>
> *IF he has not adopted THEN do nothing ELSE make her drop it*

EXPERIMENTS & RESULTS

Initiation & Experimental Setup

In all of the experiments, diffusion starts with initiation of it via making certain fraction of agents from each status adopt the innovation. There are many different ways to select these early adopters. However, in this study we will only report the results of the experiments with the selection of most sociable agents -the ones with maximum number of friends- as early adopters. Note that this choice is in favor of adoption. Furthermore, optimistically–after having many trials-, it is decided that 10% of the population will adopt the innovation at time zero with an equal number of agents from both statuses. Moreover, it is seen that the meaningful cases mostly occur when adoption threshold is around 30%. If it is lower this value, than the diffusion is too easy that having a differentiating observation is impossible. If it is higher, then the diffusion turns out to be too hard that again, drawing conclusions is impossible.

There are two important points about the results reported in this paper. Firstly, they are the most interesting ones among many others that are obtained via extensive experimentation. A range of values are tried out for parameters such as average degree, number of nodes, and many other network parameters. Secondly, the experiments are replicated on significantly many different networks and averages of the results are reported.

Experimental Design

In order to reveal the impact of homophily, we will try different "happiness thresholds", i.e. *%-similar-wanted,* to obtain different levels of homophily across the network. We will keep track of *%-similar-total,* as the network-wide (overall) homophily level. As we increase *%-similar-wanted*, we will get a more homophilious network and it is visible in *Figure 2*. Note that the homophily level always turns out to be more than what is wanted. It is clear when the three lines are compared with the reference line. This observation suggests us that when agents interact according to their individual preferences, the global level of homophily becomes more than agents' intended levels. In other words, even a small homophily tendency can reinforce itself and lead to more homophily than desired.

Results

The diffusion dynamics for different levels of homophily can be seen in *Figure 3* and *Figure 4.* In these plots y-axis refers to the fraction of population that is adopted at the final equilibrium level. Note that, to decrease the confounding effect of randomness, all the values in all plots are averages of results of 500 simulation runs per case (e.g. in *Figure 4* 100 nodes network, *%-adopted* is 26.5%, that is the average of 500 simulation runs where *%-similar-wanted* is 85% and average degree is 8 in each of these). After investigation of these plots, four important observations become noteworthy:

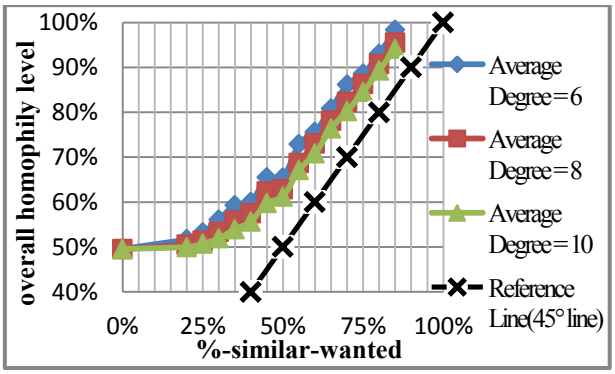


Figure 2: The homophily level changes proportional to parameter %-similar-wanted.

Firstly, both in *Figure 3* and *Figure 4,* it is seen that for low values of homophily, the adoption fractions are higher. However, after some point, as homophily level increases the adoption decreases. This macro-behavior is valid for all networks considered, 100 or 200 nodes networks and 6, 8, or 10 average degrees networks. With a low resolution view of these plots, it can be said relatively small degrees of homophily is in favor of homophily. Nevertheless, after a certain amount, it leads to decrease in the adoption.

Secondly, comparing 100 nodes network and 200 nodes network, it is obvious that the adoption is higher in 200 nodes network. This makes sense since the initial number of adopted agents is higher than 100 nodes network, allowing a higher adoption rate initially.

Thirdly, comparing the different average degree networks, it is clearly observed that as average degree increases, the adoption decreases. In other words, as agents becomes more sociable and have more friends, the overall adoption decreases. This might seem counterintuitive due to the fact that having more friends, the probability of having a friend similar to you and who adopted the innovation is higher. On the other hand, the social persuasion becomes harder since the threshold of adopting the innovation becomes higher for an agent with more friends. It turns out that the latter mechanism is more dominant.

Fourthly, although macro-behavior is understandable, micro-behavior is confusing and in some cases is in contradiction with the macro-behavior. Specifically, after the adoption makes a peak –the trend changes from "as homophily increases, adoption also increases" to "as homophily increases, adoption also decreases"- there are some local minima and maxima along the plots. Apparently, the behavior –local minima and maxima along the plot- is persistent, it is independent of number of nodes (the persistancy of the patterns can be discerned more easily in *Figure 4* since the data points lay on the same x-coordinates on the plots).

Furthermore, almost the same kind of behavior repeats itself in different average degree networks.

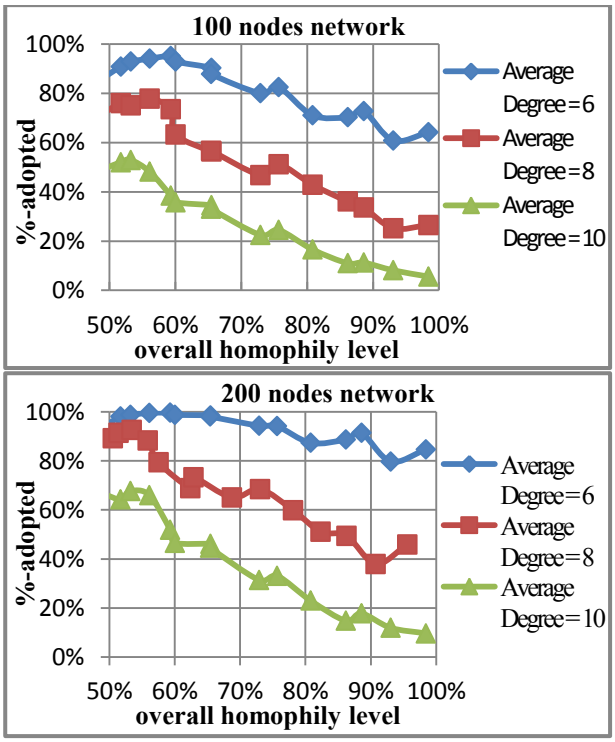


Figure 3: The plots of adoption percentages versus homophily level in the social network

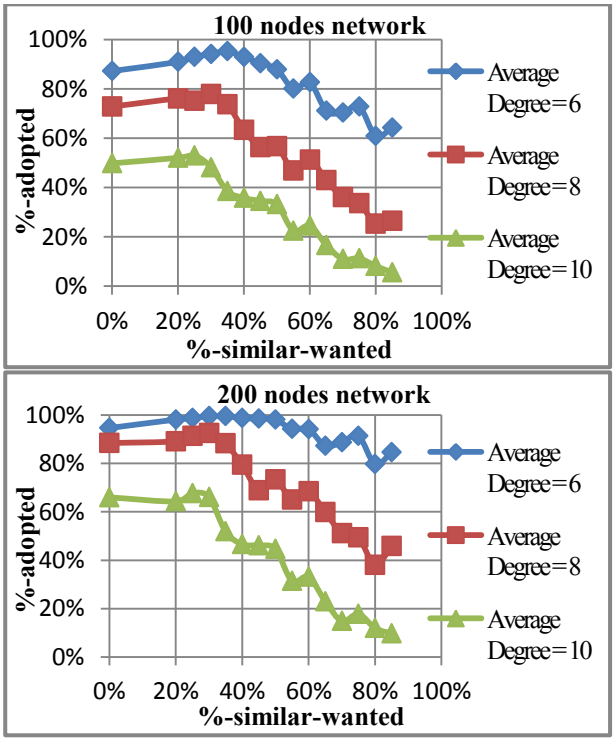


Figure 4: The plots of adoption percentages versus "happiness" thresholds

DISCUSSIONS

In this section, we will try to elaborate on the key remarks from the previous section. First, why initial increases in homophily reinforce the diffusion will be elucidated. Secondly, the negative impact of homophily will be analyzed.

The social networks studied in this paper are actually the amalgamation of two networks: the networks of the green agents and the networks of the red agents. Putting one on the top of the other gives us the same network if we were to divide it into two since diffusion takes place solely over homophilious ties. Hence, we will focus on one of the networks in isolation. For instance, *Figure 5* has two examples of the networks of the green agents. It is seen that the network with no homophily at all -the network on the left- has three disconnected groups whereas the mixed network is connected (in the homophilous network evolution routine we make sure that the emerging network is connected). Having disconnected groups, it may be impossible to reach full adoption if the initial adopters are not present in some disconnected groups. However, a slight increase in the level of homophily, the networks of green agents usually turns out to be connected and in a connected network full adoption is much more likely.

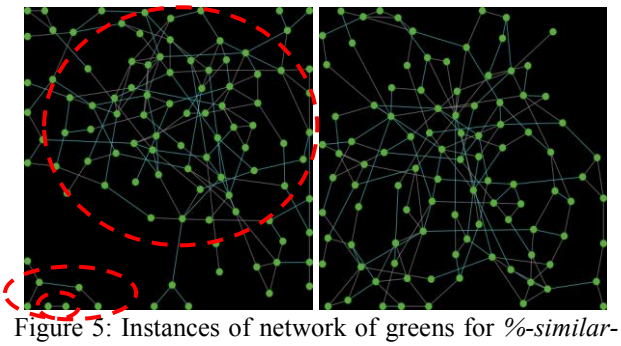


Figure 5: Instances of network of greens for *%-similar-wanted* values of 0% and 30%, respectively.

Another supporting argument to the above explanation can be found in *Figure 6* where we see the plot of average number of "weak components". A weakly connected component is simply a group of nodes where there is a path from each node to every other node. It is seen that a slight increase in the homophily level is followed by the number of weak components decreasing to one, saying that the network is connected.

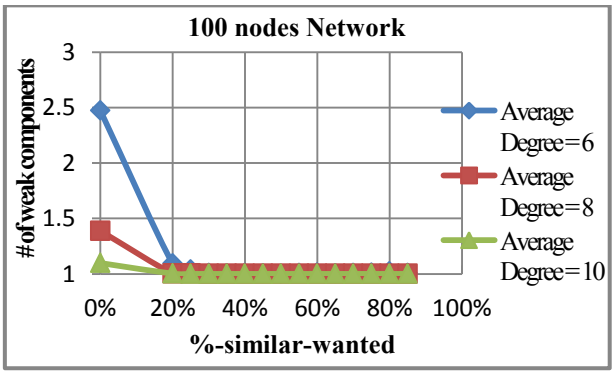


Figure 6: The plots of average number of weak components in the networks of green agents.

To explain the negative effect of homophily, we are inspired from the explanation of the impact of average degrees on the adoption. In *Figure 7*, we see that as the homophily increases in the network of the green agents, the average degree increases as well, leading to decrease in adoptions via making social persuasion to adopt harder.

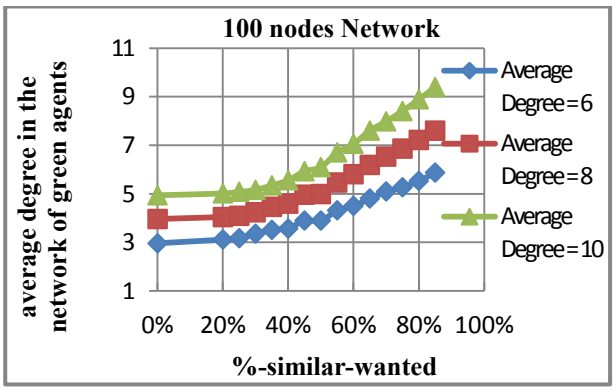


Figure 7: The plots of average degrees in the networks of green agents.

These explanations are still incomplete to explain the macro behavior of the diffusion with respect to homophily. When the network evolves homophiliously, the topology and all statistics of the network actually changes. *Figure 8* has the most important network statistics, average path length and clustering coefficients. The values are the averages of the results of 500 simulation runs. As it is clear, the trends in all are consistent.

These changes in average path length and clustering coefficient also have impacts on diffusion. However, their effect cannot be understood easily since they are distorted with the impact of homophily. To see the impact of the changes in average path length and clustering coefficient, randomly changing the statuses (i.e. colors) of agents after the network evolve homophiliously can help. An instance of such a trial can be found in *Figure 9*. First, the network evolves with respect to each *%-similar-wanted* parameter. After the network reaching the equilibrium, half of the agents' colors are swapped and the homophily levels (i.e. *%-similar-total*) are stabilized to 50%. Nevertheless, the topology of the network has already changed in the same way as in previous cases. Hence, the clustering coefficient and average path length values follow the same trends as in *Figure 8*.

From *Figure 9*, it is seen that the impact of the changes in clustering coefficient and average path length is illegible. Because, in the network where average degree is 6, the adoption decreases as *%-similar-wanted* increases whereas in the network where average degree is 10, the adoption increases and then decreases.

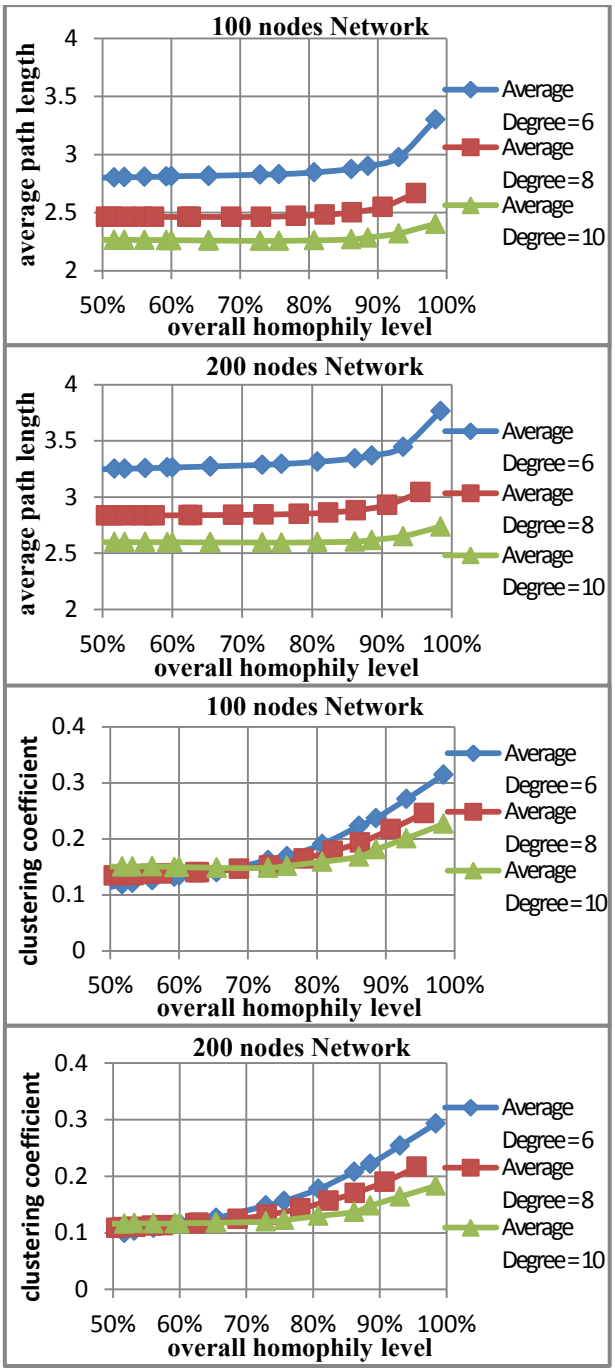


Figure 8: The plots of average path length and clustering coefficients in the homophilious networks.

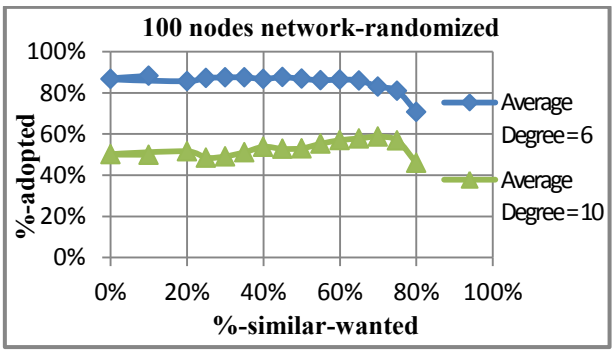


Figure 9: The plots of %-adopted in the randomized networks where %-similar-total is 50%.

CONCLUSION & FUTURE RESEARCH

In this study, it is seen that homophily reinforces itself and always leads to higher values than desired. The model presented here is capable of capturing segregation dynamics over social networks. Moreover, it is shown that small increases in homophily are in favor of diffusion whereas large increases affect negatively. It is also shown that increasing average degree of a network does not necessarily reinforce the diffusion. On the contrary, making social persuasion harder, it negatively affects the overall diffusion patterns.

The shortfall of this study, which is a future research topic, is to explain the persistent micro-level behaviors that are the local minima and maxima in the adoption plots. Although many network statistics such as clustering coefficient, average path length, average closeness-centrality, average betweenness centrality, variance of degree distributions, number of weak component, number of maximal cliques and biggest maximal cliques, and number of bi-components are tried to explain the trend, we could not come up with an explanation that elucidates the ups and downs in the decreasing part of adoption curves.

Many other future research directions can proliferate from this study and some of the promising ones are as follows: The co-evolution of the individual behavior and the network is an exciting direction. For instance, the adopted innovation can create a value homophily and during the diffusion process and accordingly, the network structure can change via new tie formations and dissolutions. However, this seems challenging since allowing the network to change during the diffusion process makes it very hard to come up with causal explanation to the observed outcome. Furthermore, having unequal number of agents in greens and reds can be interesting, addressing different societies. Lastly, different selection strategies for early adopters can yield significantly different dynamics of diffusion and hence, should be explored.

REFERENCES

Centola, Damon. 2010. "The spread of behavior in an online social network experiment." *Science (New York, N.Y.)* 329(5996):1194–97.

Centola, Damon. 2011. "An experimental study of homophily in the adoption of health behavior." *Science (New York, N.Y.)* 334(6060):1269–72.

Centola, Damon, Robb Willer, and Michael Macy. 2005. "The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms." *American Journal of Sociology* 110(4):1009–40.

Choudhury, Munmun De, H. Sundaram, and A. John. 2010. "birds of a feather": Does user homophily impact information diffusion in social media." *Arxiv preprint* 1.

Christakis, Nicholas a, and James H. Fowler. 2007. "The spread of obesity in a large social network over 32 years." *The New England journal of medicine* 357(4):370–79.

DiMaggio, Paul, and Filiz Garip. 2012. "Network Effects and Social Inequality." *Annual Review of Sociology* 38(1):93.

Epstein, Joshua M., and Robert L. Axtell. 1996. *Growing Artificial Societies: Social Science From the Bottom Up*. Washington DC: Brookings Institution Press.

Garip, F., and P. DiMaggio. 2011. "How Network Externalities Can Exacerbate Intergroup Inequality." *American Journal of Sociology* 116(6).

Gilbert, Nigel, and Klaus G. Troitzsch. 2005. *Simulation for the Social Scientist*. 2nd ed. New York: Open University.

Granovetter, Mark. 1978. "Threshold models of collective behavior." *American journal of sociology* 83(6):1420–43.

Jackson, MO. 2008. *Social and economic networks*. Princeton University Press P. 19.

Jackson, MO, and D. Lopez-Pintado. 2011. "Diffusion and contagion in networks with heterogeneous agents and homophily." *Available at SSRN 1950476* (2001):1–18.

Lazarsfeld, PF, and RK Merton. 1954. "Friendship as a social process: a substantive and methodological analysis." Pp. 18–66 in *Freedom and Control in Modern Society*, edited by M Berger. New York.

Marsden, Peter. 1987. "Core discussions networks of Americans." *American Journal of Sociology* (52):122.

McPherson, M., L. Smith-Lovin, and JM Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual review of sociology* 27(2001):415–44.

Railsback, Steven F., and Volker Grimm. 2011. *Agent-based and Individual-based Modeling: A Practical Introduction*. Princeton University Press.

Rogers, EM. 2003. *Diffusion of innovations*. 5th ed. New York: Free Press Pp. 305–8.

Schelling, TC. 1971. "Dynamic models of segregation†." *Journal of mathematical sociology* 1(2):143–86.

Wilensky, U. 2005. "Net Logo Small Worlds model."

AUTHOR BIOGRAPHIES

MUSTAFA YAVAS received his B.S. in Industrial Engineering from Boğaziçi University. He is now pursuing M.S. in Industrial Engineering from Boğaziçi University where he is also a research assistant at Socio-Economic System Dynamics Research Group (SESDYN). His research interests are system dynamics and agent-based modeling methods, and diffusions over social networks, economic inequality, and economics of population aging.

GÖNENÇ YÜCEL received his B.S. and M.S. degrees in Industrial Engineering from Boğaziçi University in 2000 and 2004. After earning his PhD degree in Policy Analysis from Delft University of Technology, he joined Boğaziçi University Industrial Engineering Department as an assistant professor. In general, Gönenç has been focusing on simulation methodology, and simulation-supported policy analysis in his research, utilizing agent-based, as well as system dynamics models. For more information; http://www.gyucel.net/Personal/Academic.html