# Who Shares Fake News in Online Social Networks?

An Agent-based Model of Different Personality Models and Behaviors in Social Networks.

Laura Burbach
burbach@comm.rwth-aachen.de
Human Computer Interaction Center
Aachen, Germany

Patrick Halbach
halbach@comm.rwth-aachen.de
Human Computer Interaction Center
Aachen, Germany

Martina Ziefle
ziefle@comm.rwth-aachen.de
Human Computer Interaction Center
Aachen, Germany

André Calero Valdez
calero-valdez@comm.rwth-aachen.de
Human Computer Interaction Center
Aachen, Germany

## ABSTRACT

Today more and more people use social networks and so the differences in personalities of users become more diversified. The same holds true for available news content. To test if regular news and fake news are distributed similarly and to what extent this depends on the personality and behavior of individuals, we conducted a mixed-method study. Through an online questionnaire we measured personality traits of individuals in social networks, how they behave, and how they are connected to each other. Using this data, we developed an agent-based model of an online social network. Using our model, an average of 92% of regular news and 98% of fake news were disseminated to the whole network. Network density turned out to be more important for dissemination than the differences in personality and behavior of individuals. Thus the spread of fake news can not only be addressed by focusing on the personality of individual users and their associated behavior. Systemic approaches—integrating both human and algorithm—must be considered to effectively combat fake news.

## CCS CONCEPTS

• **Networks → Online social networks**; • **Social and professional topics → User characteristics**; • **Computing methodologies →** *Agent / discrete models*; • **Information systems →** Personalization.

## KEYWORDS

Agent-based modeling, social simulation, fake news, online social networks, opinion forming, personality models, mixed-method study

**ACM Reference Format:**
Laura Burbach, Patrick Halbach, Martina Ziefle, and André Calero Valdez. 2019. Who Shares Fake News in Online Social Networks? : An Agent-based Model of Different Personality Models and Behaviors in Social Networks.. In *27th Conference on User Modeling, Adaptation and Personalization (UMAP '19), June 9–12, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3320435.3320456


ACM ISBN 978-1-4503-6021-0/19/06…$15.00
https://doi.org/10.1145/3320435.3320456

## 1 INTRODUCTION

In today's digital and complex world, it is becoming increasingly difficult for users of online social networks to distinguish between regular news and deliberately fake news [46]. Deciding which news are credible and which are not, requires more than human abilities and a sense of intuition. Technical skills are necessary and using digital support increases users capabilities to detect fake news [6], such as fact checking sites, which have become known in recent years. The term fake news has become well known since President Donald Trump has used this term frequently. But besides his complaints, fake news can become a threat to individuals and democracies [46].

Fake news spread through social media, where personalization algorithms influence who sees which content based on individual preferences [22, 37]. Personalization can threaten political discourse by limiting the diversity of opinions that are presented to users [30, 32], while the personalization processes itself remains hidden [42]. However, there are studies that show that online social networks expose users to more diverse opinions [17].

Fake news have not only been spread by algorithms, but foremost by people. In social media, every user can become a publisher or contribute to spreading news. The user can promote the spread easily, by marking contributions with likes, comments, or shares. All these actions increase the likelihood that the contacts in the users' network see the contribution as well. In addition, in online social networks the "circle of acquaintances" is larger than in offline life. Thus the contributions can be read by a plethora of people. Therefore, it should be critically reflected that people tend to agree with the opinions of their friends, regardless of facts [49].

In this paper, we examine how individuals use social networks online. Classical empirical methods are not suitable for the investigation of such complex social phenomena, because they cannot be described by individual behavior alone. In this study, we use agent-based modeling to investigate how individuals behave in online social networks, how they are interconnected to each other, and what user characteristics are associated with spreading regular and fake news.

## 2 RELATED WORK

In order to understand the spread of regular news and fake news, we first describe how personality influences social media use and how it can be modeled using agents in our model. Next, we focus on the specifics of fake news and how to represent them in our model. We explain how simulations of complex systems can help to understand problems like fake news and how the individual and the environment (i.e., social media) should be modeled.

### 2.1 Social networks and personality

Due to the digitization of the media world, the importance of mass media such as television, radio, and news papers are losing importance for opinion formation, compared to social media [46]. Additionally, users reveal a lot of personal information in online social networks. Users reveal who they are friends with, what hobbies they have, what musical tastes they have, what political views they hold, and much more. The behavior of users in social media, such as what they share or like, can be used to derive recommendations and further to infer much more private and sensitive information [27], such as personality.

What personality is, depends largely on how it is measured. Different theories of personality yield different dimensions of description. Therefore, we moved related work on personality measurement to the methods section. In section 2.3.1 we show why the models we have chosen are suitable for an agent-based approach.

### 2.2 Fake News

The term fake news is defined differently by different authors. We consider fake news to be deliberately posted false reports on social media. Here, every user can publish not just his own posts, but also forward other posts to a large number of people, who then might follow suit [26]. However, users are often unable to evaluate the credibility of a post or website, so they possibly do not notice when a news item is fake. Users could forward fake news without knowing so. Fake news is not a new problem, but in social media, they spread faster than in traditional media [46] and have been shown to influence elections and opinion-forming [1]. Some studies have shown that on social media fake news spread more widely than regular news.

Vosoughi et al. [50] have considered the spread of 126,000 rumor cascades on Twitter. The rumors were shared 4.5 million times between 2006 and 2017. All rumors where verified using so-called fact-checking pages. Most of the time (95–98%), the fact-checking pages agreed in their conclusion. The study showed that fake news spread faster than fact-based news. The latter usually did not reach more than 1000 people, the top 1% of fake news reached between 1000 and 100,000 people. The spread of fake news was six times faster than regular news. In terms of network penetration, fake news was passed over about 19 individual jumps, whereas regular news only made ten. Vosoughi et al. concluded that fake news has a 70% higher chance of being forwarded than fact-based news [50].

Del Vicario et al. [12] also arrived at similar results. Comparing 32 Facebook pages that publish conspiracy theories and 35 pages that publish scientific news, they found that conspiracy theories infiltrated the network much deeper. While the most frequently shared post recorded 2422 jumps, the longest cascade of scientific news only made 952 jumps. At the same time, conspiracy theories remained longer in the social network [12].

#### 2.2.1 *Action against fake news.*

The recent rapid spread of fake news has led to a gradual increase in fact-checking pages. These services scan social media for news and check their sources. The influence of these sites is, however, rather limited [45].

Serrano et al. have assumed that even when users of Twitter recognize false rumors, they will not spread counter rumors. Therefore, counter arguments are hard to find in social media. Their exploratory data analysis of two rumors about Obama and Palin on Twitter confirmed these assumptions [44]. Likewise, even if users are informed of posting fake news by a fact checking page, they rarely delete it from the network. Friggeri et al. [19] have looked at how people behave when they are told they are spreading fake news. They examined 16,672 cascades and found that only about 0.15% of users deleted the shared content [19].

In addition, users do not to trust others who provide explanations and links to fact-checking pages. When knowing the other person, 73% agreed with the explanation in a study by Margolin et al. [33]. In the other cases, only 39% accepted in that it was fake news [33].

### 2.3 Simulation of complex systems

To investigate these effects, we must first consider the underlying complexity. In complex social systems, we speak of several ontological levels. Micro- and macro-scale levels interact as subsystems of systems [10]. These systems cannot be understood by their parts alone, as the overall system is more than the sum of its parts. This is where emergent behavior occurs. To understand the whole system, it is helpful to simulate the individual subsystems [16].

For this purpose, agent-based modeling can be used [7, 16]. The system-theoretical approach with agent-based models is well suited to simulate processes such as network formation and information dissemination. In agent-based modeling, rational choice models are most frequently used [20]—which of course can be debatable with respect to fake news.

The goal of agent-based modeling is not to create an exact image of reality. Instead, an agent-based model can be used to depict individual behavior, usually greatly simplified, and system behavior can be correctly observed qualitatively. However, the evaluation of these models still poses difficulties. To evaluate such a model, an independent replication, an evaluated comparison and a validation must be available [41].

#### 2.3.1 *Agents in social networks.*

From some studies [3, 15, 27] we know that personality of users is strongly related to characteristics of their social networks. Therefore, it makes sense to model the personality of agents. To understand who distributes news or not, we model the agents in the social network as correctly as possible. Agents interact with other agents and the environment in which they are located at every simulation step. The resulting stochastic processes then depend on given probabilities [43]. To simulate the agents in our model, we next take a closer look at personality traits and behaviors of users in social networks. Those findings and the results of our survey, will be incorporated in our model.

Some studies [3, 9, 23] have found that extroverted people have more contacts in social networks and comment and post more frequently. Likewise, greater levels of openness lead to more interaction through liking, posting, and joining groups. Bachrach et al. also found that conscientious users are less likely to join groups and mark fewer posts with a "like", and that agreeable or neurotic users mark posts with a "like" rather rarely [3].

Regarding the dark triad, Sumner et al. found that narcistic users have a higher number of followers [47]. While, Mi et al. showed that people with a higher self-efficacy are more involved in the expression of positive emotions (POS) [35].

*2.3.2 Models of social networks.* In addition to the agents, the simulation environment is also modeled. It is the area in which the agents are located and determines how they interact with each other [43]. If, as in this study, the environment is a social network, an artificial social network must be created. It is attempted that the artificial networks and real social networks are structurally similar.

The generation of artificial networks has been the subject of research since the 1960s [51]. So far, the basis for such models are mostly real social networks [31]. Online social networks are large networks and usually develop uncertain structures and overlapping communities. To counteract this, agent-based modelling can be used to create networks with similar properties generatively [4, 38].

The network structure is important, because the structure is needed to model how opinions spread. Structures in online social networks reflect real social belonging [53]. They are therefore an appropriate model for linking agents in agent-based models, as in our study.

## 3 METHOD

Looking at personality traits, we investigate who interacts in online social networks. We also aim to find out how users are interconnected and how personality and density of the network influence the spread of both regular news and fake news. We conducted a mixed-method approach (see Figure 1) consisting of an online questionnaire and an agent based model.

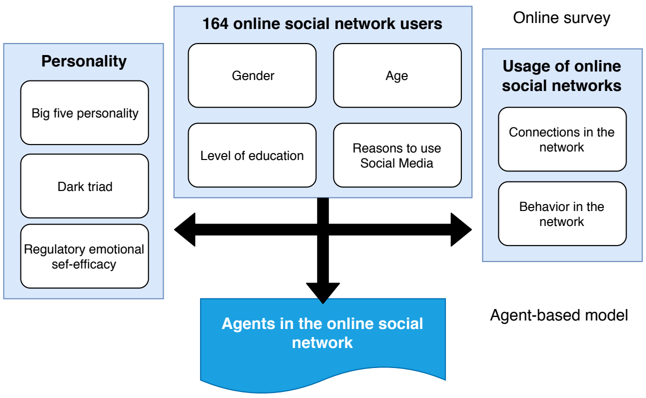


**Figure 1: Research design with the two conducted methods and the investigated variables.**

### 3.1 Online questionnaire

To find out how three different personality constructs (*Big Five, Dark Triad, Regulatory Emotional Self-Efficacy*) influence the use of social networks, we conducted an online survey in July 2018 in Germany. The online questionnaire was implemented on the platform SurveyMonkey and sent out by the authors via individual social networks (i.e., Facebook, Twitter, etc.). We used convenience sampling and, after we deleted the incomplete responses, our dataset consisted of 164 participants. The questionnaire consisted of three parts: First, demographic data was collected on *age*, *gender*, and *education level*. The second part addressed the personality of the participants. Here, we measured the *Big Five*, *Dark Triad*, and *Regulatory Emotional Self-Efficacy*. Then, we examined how participants are interconnected in social networks and how they behave online.

*3.1.1 Big Five.* To get an impression of the personality, we first used the established model of differential psychology [cf. [11, 21]]: According to the *Big Five Model*, human personality has five main dimensions: *openness*, *conscientiousness*, *extraversion*, *agreeableness*, and *neuroticism*. Many studies [14, 36, 48] have shown that people can be characterized according to these five dimensions. To keep questionnaires short, the five dimensions can be measured using short scales such as the BFI-10 scale by Rammstedt et al. or the TIPI scale by Gosling et al. [24]. We use the BFI-10, as its validity for the German language has also been shown by several studies [39].

*3.1.2 Dark triad.* We used the *Dark Triad* as an additional personality model. Delroy and Wiliams developed the model in 2012 [13]. The scale has been used and cited many times, but also been criticized by some researchers as they claim it has become to expansive [34]. The *Dark Triad* measures *narcissism*, *machiavellianism*, and *psychopathy* and how they are related. Here, too, we have used a short scale. The *insidious nine* was developed by Küfner et al. and contains three statements for each of the three constructs [29].

*3.1.3 Regulatory Emotional Self-Efficacy.* Thirdly, we measured the *Regulatory Emotional Self-Efficacy scale* (*RESE*) by Caprara et al. [8]. It consists of two dimensions, one of which is further divided into two sub-dimensions. These measure self-efficacy beliefs related to emotion regulation. The first dimension is self-efficacy in the expression of positive emotions (*POS*) and measures how individuals perceive their ability to express positive emotions such as joy, enthusiasm, and pride. The second dimension measures the perceived ability to overcome negative emotions such as frustration, anger, and despair (*NEG*). The second dimension is divided into self-efficacy in managing despondency/distress (*DES*) and self-efficacy in managing anger/irritation (*ANG*) [8, 25]. We measured *POS* with four questions and *DES* and *ANG* with three questions each.

*3.1.4 Connections in the network.* In the third section, we first looked at the connections of participants in social networks. For this, we asked how many individuals and how many institutions or people from public life the participants are connected to in all their networks (such as Twitter, Facebook and Instagram). Possible answers for both questions were *less than 50*, *50 to 100*, *100 to 300*, *300 to 500*, *500 to 1000*, and *more than 1000*.

*3.1.5 Behavior in the network.* We then asked them how often they interact with posts through liking, commenting on, or sharing them on a nine-level response scale of *never* up to *30 times a day*. If not stated otherwise [1], we measured agreement with variables

[1] except for *gender*, *age* and *education level*

on a six-point Likert scale[2] (1 - disagree completely, 6 - completely agree).

## 3.2 Agent based model

Based on the results of this survey, we have developed an agent based simulation, we called the *dissemination model*, aiming to show how messages are sent or forwarded in social media. We used the multi-agent programming language Netlogo in version 6.0.1, which was developed by Uri Wilensky [52]. One benefit of NetLogo is that it provides both user interface for designing the agent based model and testing individual simulations, as well as a batch mode to run several hundreds of simulations.

*3.2.1 Dissemination model.* When the simulation is started, a random agent sends a message to the network. This agent is referred to as the *source agent*. When agents receive the message, they decide whether to forward it or not. For visible inspection, agents who forward the message turn green and agents who received but did not forward the message turn red. The simulation stops when the message is no longer forwarded and all agents who received the message have decided on forwarding the message. Agents who have not received a message remain black.

*3.2.2 Before the dissemination model starts.* As in real life, the agents in our model have different personalities. The personality assigned to the agents is designed according to the data collected in the online questionnaire, as well as their friend count, associated institutions, their *liking*, *commenting*, and *sharing* behavior (see section 4.1.4). Before the simulation starts, the agents are linked according to an algorithm described in section 4.1.5.

The decision of an agent to forward the message or not depends on his personality and behavior in social media. To decide whether to pass on the message or not, we have used our insights on how personality traits relate to usage behavior in online social networks from the results of our survey (see section 4.1.4). We modeled this as a linear regression. The regression coefficient of each factor is used to model the strength of the relationships: The greater it is, the stronger the influence on the agent's decision in the model. The decision is made by chance, but the probability is influenced positively or negatively by the inclusion of these factors. This allows to model the behavior stochastically in accordance with the collected data. So we can observe how the personality of the individual agents influences which messages they forward.

# 4 RESULTS

The data was analyzed using SPSS and Netlogo. We first present the findings of the online survey and then the findings of the *dissemination model*, based on the survey results.

To report the statistical significance in tables we added asterisks to correlation coefficients. One asterisk (*) means a level of significance of $p < .05$, two (**) of $p < .01$ and three (***) of $p < .001$, respectively.

## 4.1 Results of the online questionnaire

Before we take a closer look at the results, we first describe the sample of the online questionnaire.

[2]We chose six levels to stay in line with previous research using these items.

*4.1.1 Sample.* A total of 164 participants took part in the questionnaire. Of these, 99 (60%) were female and 65 (40%) male. On average, our participants are young ($M$ = 28.05, $SD$ = 9.54) and well educated: 81 (49%) have a university degree, 56 (34%) a university entrance qualification, 14 (9%) a vocational baccalaureate diploma, 12 (7%) a completed vocational training, and one participant a secondary school leaving certificate.

*4.1.2 Personality.* Looking at the personality of the participants, most of the values of the *Big Five* factors are dispersed around the scale value 4.0 (see Table 1). Participants are least *neurotic* and most *open*. Table 1 shows further that the mean values for the *Dark Triad* are all below the scale mean of 3.5. This means, that participants did not achieve high scores for these scales. In contrast, slightly higher values were obtained for *DES* and *ANG* and significantly higher values for *POS* (see Table 1).

*4.1.3 Social network usage.* The results on the use of social networks are presented here. Based on these, the links of the agents in the social network and whether they forward messages or not in the *dissemination model* will be realized later.

First, we look at the links of the agents. For them we asked the participants how many friends and how many institutions they are connected with in social media. As Figure 2 shows, the participants are associated with more individuals than institutions. While most (74%) are associated with *less than 50* institutions, most (64%) are associated with *100 to 500* individuals and only a few (5%) are associated with *more than 1000*.

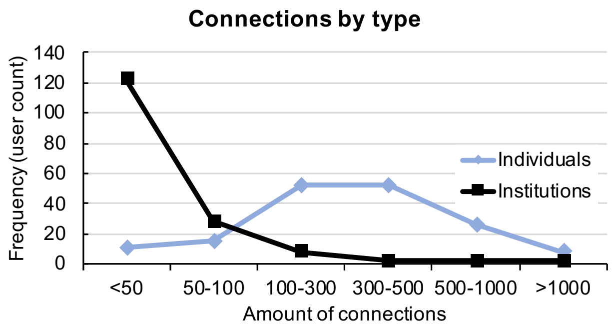


**Figure 2: Number of links to friends and institutions**

To find out whether participants tend to spread news or not, we asked how often they interacted with posts through liking, commenting on, or sharing. Figure 3 shows, that most participants never use the options "comment" (37%) and "share" (40%); and if they do, only *once a month* (c: 38%, s: 41%). Liking on the other hand is used mostly several times a week (24%).

*4.1.4 Personality and social network usage.* Since we want to implement both personality and behavior of participants in our *dissemination model*, we first show correlations between personality traits and usage of social networks. Second, using linear regressions, we show how the personality influences actual behavior.

| Dimension | Descriptives | *extr.* | *agre.* | *cons.* | *neur.* | *opn.* | *mach.* | *psy.* | *nrc.* | *POS* | *DES* | *ANG* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *extraversion* | $M = 4.02$, $SD = 1.18$ | — | .184* | | -.236** | | | | .227** | .458** | .251** | |
| *agreeableness* | $M = 3.81$, $SD = 0.89$ | | — | | | | -.242** | -.202** | | .179* | | .186* |
| *conscien.* | $M = 4.14$, $SD = 0.99$ | | | — | | | | | | .195* | | |
| *neuroticism* | $M = 3.43$, $SD = 1.12$ | | | | — | | | -.308** | | | -.554** | -.235** |
| *openness* | $M = 4.52$, $SD = 1.07$ | | | | | — | | | .210** | | | |
| *machiavel.* | $M = 3.22$, $SD = 1.00$ | | | | | | — | .334** | .293** | | | |
| *psychopathy* | $M = 2.06$, $SD = 0.84$ | | | | | | | — | | | .184* | |
| *narcissism* | $M = 3.37$, $SD = 1.08$ | | | | | | | | — | | | |
| *POS* | $M = 4.69$, $SD = 0.8$ | | | | | | | | | — | | |
| *DES* | $M = 3.49$, $SD = 0.98$ | | | | | | | | | | — | .230** |
| *ANG* | $M = 3.72$, $SD = 1.00$ | | | | | | | | | | | — |

**Table 1: Descriptives and correlations of the three personality models and traits. Only significant correlations are shown. 1.: *Big Five*, 2.:*Dark Triad*, 3.: *Regulatory Emotional Self-Efficacy***

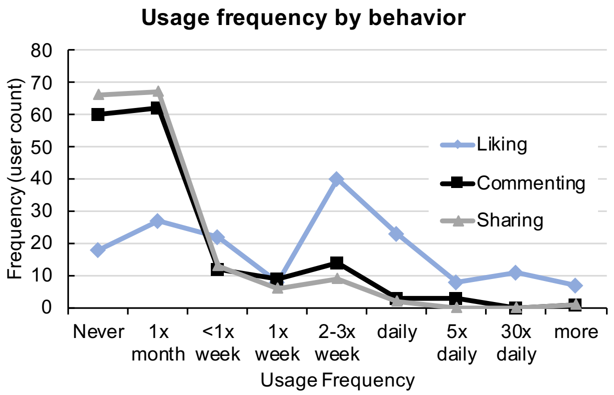


**Figure 3: Frequency of liking, commenting and sharing posts in facebook**

| metrics of social network use | correlation results |
|---|---|
| **friends** and extraversion: | $r_s(164) = .158$, $p = .044$ |
| **friends** and narcissism: | $r_s(164) = .300$, $p < .001$ |
| **institutions** and agreeabl.: | $r_s(164) = -.159$, $p = .042$ |
| **institutions** and conscien.: | $r_p(164) = -.171$,$p = .029$ |
| **institutions** and POS: | $r_s(164) = -.208$, $p = .007$ |
| **likeness** and conscien.: | $r_s(164) = -.260$, $p = .001$ |
| **likeness** and DES: | $r_s(164) = -.205$, $p = .009$ |
| **commenting** and conscien.: | $r_p(164) = -.206$, $p = .008$ |
| **sharing** and openness: | $r_s(164) = .173$, $p = .027$ |
| **sharing** and conscien.: | $r_p(164) = -.163$, $p = .037$ |

**Table 2: Correlation of the three personality and online social network usage**

| | Coeff. *B* | *SEB* | stand. $\beta$ | *p* |
|---|---|---|---|---|
| Constant: | 2.360 | .295 | | < .001 |
| ***extraversion:*** | .354 | .083 | .317 | < .001 |
| Regression $r^2 = .095$ | | | | |

**Table 3: Linear model of two predictors for the dependent variable *number of friends***

As can be seen in Table 2, people with higher *extraversion* and *narcissism* seem to have more friends on social networks. In addition, persons with higher *self-efficacy in the expression of positive emotions* (*POS*) like fewer institutions. Users who have higher scores in *agreeableness*, and *conscientiousness* and a higher *self-efficacy in dealing with despair* (*DES*) tend to leave fewer posts on social networks. In contrast, more open people are more likely to share posts.

The *number of friends* in social networks correlates positively with *extraversion* and *narcissism* (see Table 2). That is why we used a stepwise linear regression with *extraversion* and *narcissism* as independent variables and the *number of friends* as the dependent variable. The result (see Table 3) shows that a person's narcissism mostly influences how many friends they have in social networks. The final model indicates that if *narcissism* increases by one point, the *number of connected individuals* increases by 0.35 points ($F(1, 162) = 18.06$, $p < .001$, $n = 164$; $r^2 = 0.095$).

We calculated another stepwise linear regression for the *number of institutions* as dependent variable and the variables that correlated significantly (see Table 2) with the institutions (*agreeableness*, *conscienciousness*, and *POS*) as independent variables. As Table 4 shows, we have found two different models, but the second model explains more variance. This model shows, that the *number of institutions* is negatively influenced by the *self-efficacy in the expression of positive emotions* (*POS*) by 0.26 and by the *conscientiousness* by 0.15 ($F(2, 161) = 8.49$, $p < .001$, $n = 164$, $r^2 = 0.095$).

Since we consider the distribution of news in our *dissemination model*, we also measure how personality traits influence whether a user likes, shares, or comments on a post. Therefore, we have calculated step-wise linear regressions for the dependent variables *liking*, *sharing*, and *commenting*. We have used *conscientiousness* as an independent variable in all three step-wise linear regressions, since *liking*, *commenting*, and *sharing* all correlate significantly with *conscientiousness* (see Table 2). We also included *DES* as an

| | Coeff. *B* | *SEB* | stand. $\beta$ | *p* |
|---|---|---|---|---|
| Constant: | 2.663 | .372 | | < .001 |
| ***POS***: | -.266 | .078 | -.259 | .001 |
| Constant: | 3.286 | .461 | | < .001 |
| ***POS***: | -.264 | .077 | -257 | .001 |
| ***conscien.***: | -.153 | .068 | -.168 | .026 |
| Regression $r^2$ = .067 for Step 1; $r^2$ = .095 for Step 2 ($p < .05$) | | | | |

**Table 4: Linear model of two predictors for the dependent variable *number of institutions***

independent variable for *liking* and *openness* as an independent variable for *sharing* (see Table 2). The results show that all dependent variables are negatively influenced by *conscientiousness*. For *liking* the better of two models explains 9.8% of the variance. The model ($F(2, 161) = 9.88$, $p < .001$) shows that in addition to *conscientiousness* ($b = -.576$, $p = .001$), *DES* ($b = -.391$, $p = .025$, $n = 164$) also has a significant negative influence on *liking*. For *commenting* we only calculated a regression with *conscientiousness* as a predictor. The model ($F(2, 161) = 7.22$, $p = .008$, $n = 164$, $r^2 = .037$) showed that if conscientiousness increases by one point, the participants comment by -0.32 less. Finally, we calculated a step-wise linear regression with *sharing* as dependent variable and *conscientiousness* and *openness* as independent variables and found only one model ($F(1, 162) = 4.44$, $p = .037$, $n = 164$). It explained only 2.1% of the variance and only *conscientiousness* showed a significant influence ($b = -.210$, $p = .037$).

*4.1.5 Integration into the dissemination model.* The linear regression results described in chapter 4.1.4 form the basis of our *dissemination model*. The decision of the agents whether to forward a message or not is weighted by the regression coefficients. Here, we describe how these results are integrated into our model. The heart of the simulation can be summarized in two formulas. The first formula indicates how many friends or connected individuals the agents have (see 1). For this we use the results of the first regression analysis with the *number of friends* as dependent variable (see Table 3).

**Snippet 1: Network creation**

```
ask turtles [
    let fr_count ( 2.36 + 0.354 * narcissism )
    if random ( density )  > density - log
         fr_count 2 [
        create-links-with n-of 1 other turtles
    ]
]
```

Snippet 1 shows the code from NetLogo where `fr_count` stands for the number of friends. The code shows that the narcissism of each agent is multiplied by the regression coefficient *b* (i.e., 0.334) and added to the regression constant, resulting in the number of friends (`fr_count`).

If a random value is higher than a threshold, new connections are inserted. The threshold is calculated by subtracting the number of friends (`fr_count`) from a `density` value. This results in agents with a higher narcissism value making the network more interconnected.

**Snippet 2: News spread**

```
to Check
  ask turtles with [ any? link-neighbors with
       [color = green] and check = false ]  [
    let pers ((2.663 + -0.266 * POS) + (7.036 +
         -0.576 * conscientiousness + -0.391 *
         DES))
        ifelse Fakenews * random 5 > ( 5 - log
             pers 2)
    [
      set color green
      set check true
    ]
    [
      set color red
      set check true
    ]
  ]
```

Next, we need to determine whether an agent forwards a message or not. For this, we have created an algorithm (see Snippet 2) based on the results of the other four regression analyses. According to this code, the agents check whether their neighbors have received the message and want to forward it. As mentioned in section 3.2.1, agents who forward the message turn green. If an agent now detects green agents in its environment, it creates a `pers`-variable. The `pers`-variable is created based on the linear regression results on the *number of institutions* and *liking*. The `pers`-variable is then compared with a random number between 0 and 5. If this random number is greater than 5 minus the log-2 of the `pers`-variable, the agent forwards the message.

When we put numbers in the formula, the highest damping occurs if all personality scales assume the value 6 and a value of 1.202 results. At 3.5 (the scale center) the formula resolves to 2.429 and we speak of an average attenuation. At 1 we get the lowest attenuation with the formula value 3.082. We have set the comparison between `pers` and the random number so that at an average attenuation the probability that the message will be forwarded is 50:50.

In addition, we have implemented a further factor in the model. For each run either regular news or fake news is sent. When fake news is sent, the random number is multiplied by the variable `Fakenews` with the value 1.7, because in section 2.2 it was shown that fake news have a 70% higher probability to be forwarded.

*4.1.6 Simulation procedure.* We have run 3030 simulations with six replications (total 18.180 runs). In these 3.030 runs, 101 randomly generated networks were created and a message was sent 30 times. The 101 networks were used identically in 30 simulations, but we varied one aspect: the *source agent* was randomized 30 times. This means that the initial message was sent 30 times by a different agent. We varied the number of agents between 50, 300, and 1000. We also distinguished between regular news and fake news. The number of simulations was chosen as as a good representation of both coverage of the input factor space and probabilistic replications.

*4.1.7 Measurements.* We measured how deep the message was spread through the networks with 50, 300, and 1000 agents. Here, we determined how many agents received the message and forwarded it (*spreaders*), how many received it but did not forward it (*dead ends*), and how many did not receive it at all (*unreached*). Secondly, we measured how much the agents in the network are connected to each other. Here, we have observed how many neighbors the *source agent* has and how many neighbors two random agents (*agent pair*) (see chapter 4.1.7) have. This gives us a simple estimate of the network density.

For the network with 50 agents and regular messages we have shown that 11% of agents are *spreaders* ($SD = 11.14$), 17% *dead ends* (SD = 14.99) and 71% *unreached* ($SD = 25.37$). The number of *unreached* decreases with increasing number of agents in the network. For a network of 300 agents, only 26% are not reached ($SD = 29.71$), 47% ($SD = 18.87$) are *dead ends* and 27% ($SD = 11.24$) are *spreaders*. With 1000 agents, there are only 8% ($SD = 19.87$) *unreached*, 59% (SD = 12.77) *dead ends* and 33% ($SD = 7.34$) *spreaders*.

*4.1.8 Regular News vs. Fake News.* Next, we looked at the difference between regular and fake news. As Figure 4 shows, the number of *spreaders* increases with the size of the network for both regular news and fake news. The growth is much lower for fake news, which is more widespread from the start. From 50 to 300 agents, the number of *spreaders* increases by about 135%, whereas for fake news the growth is only 50%. The number of *unreached* decreases by about 64% for regular news and about 76% for fake news. For fake news there are then 54% ($SD = 13.08$) *spreaders*, 36% ($SD = 8.72$) *dead ends* and only 10% (SD = 21.08) *unreached*. With 1000 agents there are 58% ($SD = 7.17$) *spreaders*, 39% ($SD = 4.87$) *dead ends* and 3% ($SD = 11.63$) *unreached*.

We then calculated a linear regression with the mean percentage of *unreached* as a dependent variable and the mean common neighbors of the *source agent*, a random *agent pair*, and the *number of agents* in the network as independent variables. With the help of this regression analysis we have determined the influence of the network on the distribution of the message. The results show that all three independent variables have an influence on how many agents were not reached for both regular messages ($F(3, 9086) = 2272, 202$, $p < .001$, $n = 9090$, $r^2 = 0.43$) and fake news ($F(3, 9086) = 947, 282$, $p < .001$, $n = 9090$, $r^2 = 0.24$).

The results of the regression analysis for regular news and fake news in our simulation model show, that if the message to be spread is fake news and the number of common friends of agent 1 and agent 2 increases by one agent, the number of *unreached* is reduced by 0.47 percentage points. If the *source agent* receives a new neighbor, this percentage decreases by 0.34 percentage points. The number of agents per new agent has a negative impact of 0.02 percentage points. The more neighbors there are, there more users get the message.

In the case of regular news, another common neighbor of agents 1 and 2 causes the percentage of *unreached* to decrease by 0.73 percentage points. If the *source agent* gets another neighbor, the percentage of *unreached* points decreases by 0.67 percentage points. If another agent is added to the network, the percentage of *unreached* agents decreases by 0.04 percentage points.

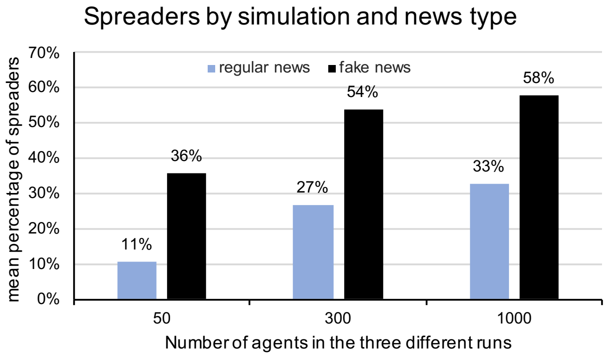


**Figure 4: Number of *spreaders* for the different network sizes**

*4.1.9 Comparison with online questionnaire.* Now, we compare the model described so far with a model in which the forwarding of the message does not depend on the personality of the agents, but only on their behavior in the network. In this model, the agents behave similarly as the survey participants state to behave in social networks.

In order to facilitate comparability, we simulated the same 101 networks for the simulation on the basis of usage only ignoring personality traits. This again gave us 3030 individual results for 50, 300 and 1000 agents as well as for regular news and fake news, which we replicated six times (18,180 runs).

A comparison of the two resulting regressions showed only minimal differences. The results of the simulation used for validation showed mean value differences of less than 10% and the standard deviation for regular news and fake news differed by only about one percent. Interestingly, personality seems to influence spread through the network to only about 10%.

## 5 DISCUSSION

Our study showed a relationship between the personality of users and their behavior in social media. This corresponds to the results of other studies [3, 9, 23, 47]. Similarly to previous studies, more narcissistic people have more friends in social networks [47]. We further confirmed a result by Bachrach [3]: higher conscientiousness makes people like less posts. In contrast to previous studies [3, 9, 23], extraversion had less impact on the behavior of online social network users. In addition, people who have a high-level of self-efficacy with positive emotions seem to interact more frequently in online social networks (liking, commenting, sharing). In our sample, only four of the eleven personality traits influenced the behavior in social networks.

Studies that have examined the relationship between personality and Facebook use so far have come to different conclusions. For example some of the findings of Amichai-Hamburger [2] contradicted the findings of Ross et al. [40]. For instance, Amichai-Hamburger found that extraversion has a positive influence on the number of Facebook friends [2], whereas Ross et al. did not [40]. The study by

Bachrach et al., which looked at a large and heterogeneous sample of 180,000 users, could be considered to be more valid in this regard. Still, they only looked at the Big Five factors and did not include other models such as the Dark Triad and the Regulatory Emotional Self-Efficacy factors. The relationship between personality traits and behavior in social networks should be re-examined in the future on the basis of larger samples that are more heterogeneous with regard to the educational level of the participants. The present level of education in the sample might have skewed the results in the direction we have reported here.

As shown in section 4.1.4 the r-square values of the regression models are rather small, which means that the four identified user factors cannot fully explain, if the users comment, like or share a message. Therefore we will consider further factors as control factors into the developed dissemination model in further studies to find out more precisely which factors have the greatest influence on how users behave.

It should be noted that the *dissemination model* is based on the information provided by the participants. Participants only expressed an assessment of their own liking, commenting, and sharing behavior as well as their own number of contacts. Therefore, the behavior in social networks and the connections between the users is only estimated. It would be desirable to verify these results in further research using actual (Facebook/Instagram/Twitter) profile features. Nevertheless, the developed model, as a simplified representation of reality, provides valuable information on how personality traits influence behavior in social networks. We have seen that a high average node degree in a network almost inevitably leads to users receiving the majority of messages. The influence of personality was rather low. This could have been caused by a relatively "normal" sample population. No users showed extreme characteristics (e.g., highly narcissistic influencer). Future simulation scenarios should focus on larger networks and thus naturally on the far ends of the normal distribution of personality.

A challenge for agent-based modeling is to develop a model that is also relevant to reality [7]. Our model is relevant as there is a multitude of fake news sent in online social networks that influences politics. Yet, the model lacks support for large-scale simulations (at least on a desktop computer) and our next steps are aimed at developing larger simulations that reflect large scale network types.

As Frias-Martinaz et al. argued, agent-based models can map individuality and randomness well, but they lose realism by mapping how people change their behavior and adapt to different situations less well. In their opinion, this is mainly due to the fact that human behavior is modelled according to data collected in surveys [18]. The dissemination model is also based on the results of an online survey. It shows well how the individual behaviour (liking, commenting, and sharing) and personality traits affect the distribution of news. So far, however, it ignores any changes or adjustments. For future studies, we plan that the personality and behavior of the agents can change and that there is no fixed endpoint in the model.

Our model was based on an Barabási-Albert model (growth and preferential attachment) with thresholds from our empirical distributions. It ignores changes in the network. Users follow and unfollow users, e.g., when they repeatedly send fake news. This is not modelled yet. Further, a more complex topic model with users individual preferences could be a good extension of this model to incorporate the fact that some users willingly share fake news when they agree with the underlying sentiment.

Another aspect, that we plan to integrate into the dissemination model is opinion leadership, so that the opinion of some agents is more important for their neighbors than the opinion of the other agents as modelled in Bianconi-Barabási style networks.

## 6 CONCLUSION AND OUTLOOK

We identified four personality characteristics that have an influence on the users behavior in online social networks. We developed an agent-based model that shows whether different users forward regular and fake news or not. The dissemination model showed that a large network usually results in almost everyone on that network being reached when a message is sent and that this phenomenon is even stronger when the content can be categorized as fake news. The developed model serves as a basis for further research in the context of message spreading and can be extended in the future.

It is not only important to push the use of fact-checking pages, but also to create awareness among users of online social networks that provide content on such platforms. Users should always scrutinize posts regarding their evidence base. Through a light-hearted click on the like-button in a high density network a cascade of sharing behavior can be triggered. Such scenarios can easily be prevented when algorithmic censoring is applied. But not all problems can be solved by algorithms and it is unsure whether users want to rely on them [5]. Here, topics like ethics, responsibility, or intimacy become important [28], because these require a sensitive and conscious decision making of humans. Solutions to this challenge can be algorithm-based and user-interface based. Providing information on the sources of fact-checking badges and personalized warning-labels could help the user to make a more informed decision.

## ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their constructive and helpful feedback. We further wish to thank Nicolas Lindstädt for designing the study and collecting the data. This research was supported by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia. Grant-Number: 005-1709-0006, project "Digitale Mündigkeit", project-number 1706dgn017.

## REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31, 2 (2017), 211–236.

[2] Yair Amichai-Hamburger and Gideon Vinitzky. 2010. Social network use and personality. *Computers in Human Behavior* 26, 6 (2010), 1289–1295.

[3] Yoram Bachrach, Michael Kosinski, Thore Graepel, Pushmeet Kohli, and David Stillwell. 2012. Personality and Patterns of Facebook Usage. In *Proceedings of the ACM Web Science Conference*. ACM, New York, NY, 36–44.

[4] Christopher L. Barrett, Richard J. Beckman, Maleq Khan, V. S. Anil Kumar, Madhav V. Marathe, Paula E. Stretz, Tridib Dutta, and Bryan Lewis. 2009. Generation and analysis of large synthetic social contact networks. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*. IEEE, New York, NY, 1003–1014. https://doi.org/10.1109/WSC.2009.5429425

[5] Laura Burbach, Johannes Nakayama, Nils Plettenberg, Martina Ziefle, and André Calero Valdez. 2018. User Preferences in Recommendation Algorithms: The Influence of User Diversity, Trust, and Product Category on Privacy Perceptions in Recommender Algorithms. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18)*. ACM, New York, NY, USA, 306–310. https://doi.org/10.1145/3240323.3240393

[6] André Calero Valdez and Martina Ziefle. 2018. Believability of News. In *Congress of the International Ergonomics Association*. Springer, Cham, 469–477.
[7] André Calero Valdez and Martina Ziefle. 2018. Human Factors in the Age of Algorithms. Understanding the Human-in-the-loop Using Agent-Based Modeling. In *Social Computing and Social Media. Technologies and Analytics*, Gabriele Meiselwitz (Ed.). Springer International Publishing, Cham, 357–371.
[8] Gian Vittorio Caprara, Laura Di Giunta, Nancy Eisenberg, Maria Gerbino, Concetta Pasto-relli, and Carlo Tramontano. 2008. Assessing regulatory emotional self-efficacy in three countries. *Psychological Assessment* 20 (2008), 227–237. https://doi.org/10.1037/1040-3590.20.3.227
[9] Kanokporn Cheevasuntorn, Bunthit Watanapa, Suree Funilkul, and Worarat Krathu. 2017. Factors influencing teenagers behavior on posting and sharing messages via Facebook. In *2017 2nd International Conference on Information Technology (INCIT)*. IEEE, New York, NY, 1–4. https://doi.org/10.1109/INCIT.2017.8257852
[10] Rosaria Conte, Nigel Gilbert, G. Bonelli, Claudio Cioffi, Guillaume Deffuant, János Kertész, Vittorio Loreto, Suzy Moat, Jean-Pierre Nadal, Angel Sánchez, A. Nowak, A. Flache, M. San Miguel, and D. Helbing. 2012. Manifesto of computational social science. *European Physical Journal-Special Topics* 214 (2012), 325.
[11] Boele De Raad. 2000. *The Big Five Personality Factors: The psycholexical approach to personality.* Hogrefe & Huber Publishers, Germany.
[12] Michael Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Science of the United States of America* 113, 3 (2016), 554–559.
[13] Paulhus Delroy and Kevin Wiliams. 2012. The Dark Triad of Persoanlity: Narcissism, Machiavellianism and Psychopathy. *Journal of Research in Personality* 36, 6 (2012), 556–563.
[14] John Digman and Naomi K. Takemoto-Chock. 1981. Factors in the natural language of personality: Reanalysis and comparison of six major studies. *Multivariate Behavioral Research* 18 (1981), 149–170.
[15] Yuxiao Dong, Yang Yang, Jie Tang, Yang Yang, and Nitesh V. Chawla. 2014. Inferring User Demographics and Social Strategies in Mobile Social Networks. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*. ACM, New York, NY, USA, 15–24. https://doi.org/10.1145/2623330.2623703
[16] Joshua M. Epstein. 2007. *Generative Social Science: Studies in Agent-Based Computational Modeling (Princeton Studies in Complexity)*. Princeton University Press, Princeton, NJ, USA.
[17] Richard Fletcher and Rasmus Kleis Nielsen. 2017. Using social media appears to diversify your news diet, not narrow it.
[18] Enrique Frias-Martinez, Graham Williamson, and Vanessa Frias-Martinez. 2011. An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 Third International Conference on Social Computing*. IEEE Computer Society, Boston, 57–64.
[19] Adrien Friggeri, Lada A. Adamic, Dean Eckles, and Justin Cheng. 2014. Rumor Cascades. In *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014.*, Eytan Adar, Paul Resnick, Munmun De Choudhury, Bernie Hogan, and Alice H. Oh (Eds.). The AAAI Press, Palo Alto, California, 100–110. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8122
[20] Nigel Gilbert. 2008. *Agent-based models. 153.* Sage, Los Angeles, CA.
[21] Lewis R. Goldberg. 1990. An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology* 59 (1990), 1216–1229.
[22] Eric Goldman. 2008. Search Engine Bias and the Demise of Search Engine Utopianism. In *Web Search: Multidisciplinary Perspectives*, Amanda Spink and Michael Zimmer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 121–133. https://doi.org/10.1007/978-3-540-75829-7_8
[23] Samuel D. Gosling, Adam A. Augustine, Simine Vazire, Nicholas Holtzman, and Sam Gaddis. 2011. Manifestations of personality in Online Social Networks. Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, behavior and social networking* 14, 9 (2011), 483–488.
[24] Samuel D. Gosling, Peter J. Rentfrow, and William B. Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 37 (2003), 504–528.
[25] Catherine Gunzenhauser, Tobias Heikamp, Maria Gerbino, Guido Alessandri, Antje Von Suchodoletz, Laura Di Giunta, and Gisela Trommsdorf. 2013. Self-Efficacy in Regulating Positive and Negative Emotions. *European Journal of Psychological Assessment* 29, 3 (2013), 197–204.
[26] Andreas M. Kaplan and Michael Haenlein. 2010. Users of the world, unite! The challenges and opportunities of Social media. *Business horizons* 53, 1 (2010), 59–68.
[27] Michael Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *PNAS* 110, 15 (2013), 5802–5805.
[28] Felicitas Kraemer, Kees Van Overveld, and Martin Peterson. 2011. Is there an ethics of algorithms? *Ethics Inf. Technol.* 13, 3 (2011), 251–260.
[29] Albrecht C.P. Küffner, Michael Dufner, and Mitja D. Back. 2015. Das dreckige Dutzend und die Niederträchtigen Neun. *Diagnostica* 61, 2 (2015), 76–91.
[30] Matthias Leese. 2014. The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. *Security Dialogue* 45, 5 (2014), 494–511.
[31] Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. 2009. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* 6, 1 (2009), cited By 499 29–123.
[32] Kevin Macnish. 2012. Unblinking eyes: The ethics of automating surveillance. *Ethics and Information Technology* 14, 2 (2012), 151–167.
[33] Drew B. Margolin, Aniko Hannak, and Ingmar Weber. 2017. Political Fact-Checking on Twitter. When Do Corrections Have an Effect? *Political Communication* 35, 2 (2017), 196–219.
[34] John W. McHoskey, William Worzel, and Christopher Syzarto. 1998. Machiavellianism and psychopathy. *Journal of Personality and Soc. Psy.* 74 (1998), 192–210.
[35] Shi Mi, Li Xupei, Zhu Tingshao, and Shi Kan. 2010. The Relationship between Regulatory Emotional Self-efficacy, Big-Five Personality and Internet Events Attitude. In *IEEE 2nd Symposium on Web Society*. IEEE, New York, NY, 61–65.
[36] Warren T. Norman. 1963. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology* 66 (1963), 574–583.
[37] Eli Pariser. 2011. *The filter bubble : What the Internet is hiding from you.* UK: Viking, London.
[38] Minh-Duc Pham, Peter Boncz, and Orri Erling. 2013. S3G2: A Scalable Structure-Correlated Social Graph Generator. In *Selected Topics in Performance Evaluation and Benchmarking*, Raghunath Nambiar and Meikel Poess (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 156–172.
[39] Beatrice Rammstedt, Christoph J. Kemper, Mira Céline Klein, Constanze Beierlein, and Anastassiya Kovaleva. 2013. A Short Scale for Assessing the Big Five Dimensions of Personality. 10 Item Big Five Inventory (BFI-10). *methods, data, analyses* 7, 2 (2013), 233–249.
[40] Craig Ross, Emily S. Orr, Mia Sisic, Jaime M. Arseneault, Mary G. Simmering, and Orr R. Robert. 2009. Personality and motivations associated with facebook use. *Computers in Human Behavior* 25, 2 (2009), 578–586.
[41] Juliette Rouchier, Claudio Cioffi, J. Gary Polhill, and Keiki Takadama. 2008. Progress in model-to-model analysis. *Journal of Atif. Societies and Soci. Simulation* 11, 2 (2008), 8.
[42] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2018. Auditing algorithms: Research methods for detecting discrimination on Internet platforms.
[43] Emilio Serrano and Carlos A. Iglesias. 2016. Validating Viral Marketing Strategies in Twitter via Agent-based Social Simulation. *Expert Syst. Appl.* 50, C (May 2016), 140–150. https://doi.org/10.1016/j.eswa.2015.12.021
[44] Emilio Serrano, Carlos Ángel Iglesias, and Mercedes Garijo. 2015. A Novel Agent-Based Rumor Spreading Model in Twitter. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion)*. ACM, New York, NY, USA, 811–814. https://doi.org/10.1145/2740908.2742466
[45] Chengcheng Shao, Pik-Mai Hui, Lei Wang, Xinwen Jiang, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. 2018. Anatomy of an online misinformation network. *PloS one* 13, 4 (2018), e0196087.
[46] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explor. Newsl.* 19, 1 (Sept. 2017), 22–36. https://doi.org/10.1145/3137597.3137600
[47] Chris Sumer, Alison Byers, Rachel Boochever, and Gregory J. Park. 2012. Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets. In *(Bd.2) 11th International Conference on Machine Learning Applications*. IEEE, New York, NY, 386–393.
[48] Ernest C. Tupes and Raymond E. Christal. 1992. Recurrent personality factors based on trait ratings. *Journal of Personlity* 60, 2 (1992), 225–51.
[49] Jason Turcotte, Chance York, Jacob Irving, Rosanne M. Scholl, and Raymond J. Pingree. 2015. News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking. *Journal of Computer-Mediated Communication* 20 (2015), 520–535.
[50] Soroush Vosoughi, Derb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.
[51] Stanley Wasserman and Philippa Pattison. 1996. Logit models and logistic regression for social networks: I. An introduction to markov graphs and p*. *Psychometrika* 61, 3 (1996), cited By 517, 401–425.
[52] Uri Wilensky. 1999. Center for Connected Learning and Computer Based Modeling. Northwestern University. Retrieved January 31, 2019 from http://ccl.northwestern.edu/netlogo
[53] Elena Zheleva, Hossam Sharara, and Lise Getoor. 2009. Co-evolution of Social and Affiliation Networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*. ACM, New York, NY, USA, 1007–1016. https://doi.org/10.1145/1557019.1557128