# Modeling Contagion of Behavior in Friendship Networks as Coordination Games

Tobias Jordan[1] and Philippe de Wilde[2] and Fernando Buarque de Lima-Neto[3]

[1] University of Kent, United Kingdom
tj202@kent.ac.uk,
[2] University of Kent, United Kingdom
P.Dewilde@kent.ac.uk,
[3] University of Pernambuco, Brazil/ University of Münster (ERCIS), Germany
fbln@ecomp.poli.br,

**Abstract.** It has been shown that humans are heavily influenced by peers when it comes to choice of behavior, norms or opinions. In order to better understand and help to predict society's behavior, it is therefore desirable to design social simulations that incorporate representations of those network aspects. We address this topic, by investigating the performance of a coordination game mechanism in representing the diffusion of behavior for distinct data sets and diverse behaviors in children and adolescent social networks. We introduce a set of quality measurements in order to assess the adequacy of our simulations and find evidence that a coordination game environment could underlie some of the diffusion processes, while other processes may not be modeled coherently as a coordination game.

## 1 Introduction

Christakis and Fowler [1] suggest that our social contacts heavily influence our decisions, opinions and and behavior. By proposing new approaches to tackle existing problems in crowd behavior, these findings may be helpful for people that aim to understand, guide or control the behavior of societies. For example, the knowledge that the political opinions of spouses strongly affect each other seems to be of high value to politicians during election periods. Physicians and other health care professionals could use the findings that adolescents affect each other's smoking, drinking and sexual-interaction behavior. Thus, the existing research on contagion of behavior and spread of information within social networks can give good hints for campaigns and political action in order to achieve a desirable behavior. Current research proves that such influence exist and is also able to point out situations where the effects are stronger or weaker. However, to adequately model the effects of contagion and information spreading for simulation and prediction models, a coherent representation is required. This is the motivation of the work presented below. Our research implements the contagion of behavior within friendship networks as a coordination game [2] [3] where single individuals within the observed systems coordinate their actions with their neighbors. We assume that individuals benefit from compliant behavior.We adapt the coordination game mechanism to two different data sets, each containing information about friendship ties, as well as time-dependent information about specific behaviors or behavioral outcomes. To evaluate the performance of the implementation of the coordination game, we simulate the coordination dynamics on the given friendship network starting with the real initial situation and subsequently compare the state of the system after the simulation with the state observable in reality after one or more time-steps.

**Synopsis.** We review the state of relevant research regarding diffusion of behavior and information in Social Networks and present the data sets used within this work in Sect. 2. Sect. 3 presents the implementation of the coordination game and its adaption to the different data sets while Sect. 4 contains experimental setup and results. We discuss the experimental results in Sect. 5 and conclude with Sect. 6.

## 2 Background

### 2.1 Diffusion processes in Social Networks

Existing research indicates that human decisions, opinions, norms and behavior are influenced by the social environment [4]. Social influence and contagion as well as spread of behavior and information through social networks has been documented in a wide range of cases [1]. For instance diffusion of voting behavior [1] and obesity [5] has been proven statistically. Moreover, cooperative behavior has been shown to be contagious, though depending on tie structure and dynamics [6] and recent studies revealed contagiousness of emotions [7]. Other behaviors do not spread like sexual orientation [8]. This indicates the existence of those effects on other individual behaviors of children and adolescents such as "commitment to school education", substance use or sport. Marques [9] reveals the huge differences between social networks of the poor and those of more wealthy people. Considering the above, this further encourages the modeling and simulation of social network effects in order to understand social phenomena and to guide political decision making. Related research has also been conducted in children and adolescent networks. For instance roles of nodes within a network of school children have already been identified [10] and diffusion of social norms and harassment behavior in adolescents school networks have been empirically studied and evidenced [11]. It then has been reasonably shown that behavior, norms information and opinions flow within social networks of adults and children. Approaches to model this diffusion come for example from the field of Social Psychology, like the concept of Social Influence Network Theory [12] from Friedkin. Another approach is the modeling as a coordination game [2] [3]. Those models may be considered as advanced threshold models [13] [14] [15] that incorporate social network structure instead of simple crowd behavior. The coordination game as implemented in [3] is characterized by the assumption that individuals benefit when their behavior matches the behavior of their neighbors in the network. Hereby a node within a network can adopt one of two behaviors $A$ or $B$. The node receives pay-off $a$ when equaling its behavior with a neighbor that adopts $behaviorA$. $b$ respectively denotes the pay-off a node receives when both, her and her neighbor adopt $behaviorB$. When choosing different behaviors, nodes receive a pay-off of 0 (other implementations may introduce negative pay-offs for non compliance).

The total pay-off for each node can accordingly be calculated as presented in 1 and 2. Here $P_i^a$ denotes the total pay-off for node $i$ from choosing behavior $A$ (respectively Behavior $B$ for $P_i^b$), $d_i$ denotes the degree of node $i$ and $n_i^a$ (same for $n_i^b$) denotes the number of neighbors of node $i$ adopting behavior $A$ (respectively Behavior $B$).

$$P_i^a = an_i^a \tag{1}$$

$$P_i^b = b(d_i - n_i^b) \tag{2}$$

This determines that the best strategy for node $i$ is to choose behavior $A$ if $n_i^a \ \geq bd_i$ and behavior B otherwise. Rearranging the inequality in 3 , we get:

$$r \geq T \text{ with } T = \frac{b}{a+b} \text{ and } r = \frac{n_i^a}{d_i} \tag{3}$$

In the absence of knowledge of the individual pay-offs $a$ and $b$, a global threshold $T$ may be found experimentally, as shown in the remainder of this paper.

### 2.2 Data

We perform our experiments on two different data sets, both contain information about adolescent friendship ties, as well as about different types of behavior.

$(i)$ The first data set stems from the study *'Determinantes do desempenho escolar na rede de ensino fundamental do Recife'* [16]. The survey was conducted by Fundação Joaquim Nabuco (FUNDAJ) in 2013, gathering data from more than 4000 pupils in public schools in the North-Eastern Brazilian city Recife. Those data contain among others the social network of the pupils and their performance in the subject maths at the beginning and at the end of the year. Children were asked to nominate their 5 best friends. In this way, a network containing 4191 students was generated. However, 573 students that did not nominate any friend within their class were removed from the data set, leading to a total number of 3618 vertices.

$(ii)$ The second data set is a selection of 50 girls from the social network data collected in the Teenage Friends and Lifestyle Study[17]. Here the friendship network, as well as behavior in sports and substance use of students from a school in Scotland were surveyed. The survey started in 1995 and continued for three years until 1997. Students were 13 years old when the study started. The study counted 160 participants of whom 129 participated during the whole study. The friendship networks were surveyed asking the pupils to name up to twelve friends. Pupils were also asked to report their behavior related to, sports, smoking as well as alcohol and cannabis consumption.

## 3 Diffusion of behavior modeled as a Coordination Game

We model the imitation of behavior of neighbors within the friendship network according to the coordination game as presented in Sect. 2. As indicated in Sect 2, we posses no information about possible pay-offs $a$ and $b$ or eventual costs of transition and hence hence aim to find the threshold $GT$ experimentally. This means that a vertex within the network changes its state over time depending on the state of its neighbors. For simplicity, the vertices may adopt one of two different states according to the investigated behavior. Hereby one state indicates that the vertex adapted behavior A, the other possible state indicates the adaption of behavior B. For each iteration, the current ratio $r_i$ is being calculated. Here $a_i$ denotes the number of neighbors of node $i$ that adapt behavior $A$ and $n_i$ denotes the total number of neighbors of node $i$.

$$r_i = \frac{a_i}{n_i} \tag{4}$$

If the perceived ratio $r_i$ is higher than the global threshold $GT$ and the state of node $i$ is B , the node changes its behavior towards Behavior A. Conversely, if $r_i$ is below $GT$ and Node $i's$ behavior is A, it changes its behavior towards B. Due to differences in data representation, the coordination game had to be implemented slightly differently for the two settings, as follows.

### 3.1 FUNDAJ

The only information available for more than one moment in time of the FUNDAJ survey is the mark of the pupils in the subject maths for the beginning and the end of the year. Although marks are not a behavior in themselves, they stem among others from individual behavior such as doing homework, paying attention, studying frequently etc. Marks are therefore considered a good indicator for the behavior *engagement at school*. They are represented as numeric values between 0 and 100. In order to differentiate between two behaviors, students are classified as *good students* or *bad students* according to their mark. Students whose mark lies below the threshold $tm$ are thereby classified as bad students and vice-versa. The setting of $tm$ defines hereby the number of *good students* (positives) and *bad students*(negatives) and hence affects heavily if nodes are predominantly connected to positives or negatives. High values for $tm$ generate large numbers of *bad students* and smaller numbers of *good students* and vice versa. The ratio $r_i$ from Equation 4 is being calculated for each student at each iteration of the simulation. If required, the mark for the next time step $m_{i+1}$ is being multiplied by the factor $1 + f$in order to alternate the state of the node:

$$m_{i+1} = m_i * (1 + f) \tag{5}$$

Parameter $GT$ sets the affinity of the nodes to change behavior. Thus, depending on the proportions of positives and negatives, it either yields a volatile or a stable system. Adaption parameter $f$ also influences the stability of the system, where volatility increases with increasing values of $f$.

### 3.2 Scottish Dataset

The Scottish data set contains information about four different behaviors, which are practicing sports, drug (cannabis) use, alcohol use and smoking behavior. Characteristic values differ slightly for the distinct behaviors, as there are for example two increments representing the intensity of sports but four increments for drug use intensity. Thus, we classified the characteristic values in order to obtain a simplified two status situation. Tab. 1 presents the characteristic values and their classification as *Behavior A*, all other values are accordingly classified as *Behavior B*. In contrast to the FUNDAJ data, the

Table 1: Classification of characteristic values for behavior

| Behavior | Characteristic values | Class. as Behavior A if: |
|---|---|---|
| Sports | 1 (non regular); 2 (regular) | $\geq 2$ |
| Drugs | 1 (non), 2 (tried once), 3 (occasional) and 4 (regular) | $\geq 2$ |
| Alcohol | 1 (non), 2 (once or twice a year), 3 (once a month), 4 (once a week) and 5 (more than once a week) | $\geq 2$ |
| Smoke | 1 (non), 2 (occasional) and 3 (more than once a week) | $\geq 2$ |

representation of behavior by discrete values required a slightly different imitation process. Hence, for the Scottish data set, if a vertex changes its state, it respectively raises the behavior value by 1 if it aims to adopt behavior A or, decreases the behavior value by 1 if it aims to adopt behavior B. Information is available for three consecutive years. Hence the starting value for each vertex in the coordination game is its behavior in year one. The quality of the simulation is measured comparing the state of the simulation after a certain number of iterations with the state of the real system after two years, here referred to as *benchmark t+1* or after three years, denominated as *benchmark t+2*. Moreover, the friendship network of the girls in the study has been surveyed for each of the three years, the study lasted. This yields the three slightly different networks $g_1$ at the first survey,$g_2$ after one year and $g_3$ after two years.This implicated for the simulation that the neighbors that a vertex considers for the calculation of its state vary for each year $t$ according to the network $g_t$ In order to incorporate those network dynamics into the simulation, we changed the network used to define the adjacent vertices of a node after completing 50% of iterations. experiments indicated that network combination of $g_1$ as representation for the friendship network in period between year 1 and year 2, and $g_2$ representing the friendship network in the period from year 2 to year 3, outperformed the results for network combination $(g_2, g_3)$. Hence, we assume that the more appropriate network combination is the former. Therefore experiments and results presented in the remainder of this paper refer to network combination $(g_1, g_2)$.

## 4 Experiments

### 4.1 Experimental Setup

Experiments were run for the two coordination game settings with varying parameters in order to find a parameter setting that leads to plausible results. As for the simulation with FUNDAJ data, the simulation was conducted with all combinations of the parameters $GT$(global threshold) and $f$(adaption parameter) for $GT, f \in [0, 0.2, 0.4, 0.6, 0.8, 1]$ and $tm$ (classification of marks) with $tm$ ∈[20,40,60,80]. For simulations with the Scottish data set the parameter $GT$ was set to values $GT \in [0.0, 0.05, 0.1, \cdots, 1.0]$.

### 4.2 Quality measurement

In order to assess the quality of the respective simulation, four distinct quality measures were applied: $(i)$ match quality, $(ii)$ ROC-curves $(iii)$ graph-based quality measures and $(iv)$ average estimation error.

$(i)$ The most intuitive measure for the simulation quality is to compare the state of each vertex $v_s$ after a certain number of simulation iterations with its state in reality $v_r$ in *benchmark t+1* or *benchmark t+2*. Hereby, we denote the case when $v_s$ = $v_r$ as $match$ and accordingly the case $v_s \neq$v$_r$ as $miss - match$. This quality measure is named $match - quality$ and denoted as $q$ for the rest of this work. The match quality $q$ of the simulation can then be assessed as in 10, where $n$ denotes the total number of vertices:

$$q = \frac{\sum_{i=1}^{n} match_i}{n} \tag{6}$$

However, for skewed attribute distributions, this measure favors estimates with high numbers of positive or respectively negative estimates and hence fails to mirror the quality of the simulation when the distribution of attributes is skewed.

$(ii)$ The ROC-metric [18] sets the number of true positives $(Recall)$ in relation to the number of false positives $(Fallout)$. $Recall$ is the ratio of correctly estimated positives values, the $true{-}positives$ and the total number of positive values $n^p$. $Fallout$ denotes the ratio between wrongly estimated positive values $false - positives$ and the total number of negative values $n^n$.

$$Recall = \frac{true - positives}{n_p} \tag{7}$$

$$Fallout = \frac{false - positives}{n_n} \tag{8}$$

The ROC-curve displays respectively $Recall$ values for each simulation on the ordinate and $Fallout$ values on the abscissa. Values above the diagonal of the graph indicate the existence of a signal and values below the diagonal may be interpreted as noise. Thus, this metric provides a clearer picture of simulation quality. Best estimates can be found mathematically maximizing the $Youden - Index$ [19] $y$ as presented in 9.

$$y = Recall - Fallout \tag{9}$$

$(iii)$ For global analysis it might not be necessary to simulate the state of each vertex correctly, as long as the system state can be predicted adequately. Thus, as third quality measure, *behavior distribution in friendship-patterns* was implemented. Hereby we define friendship patterns in the network using a modified version of NEGOPY [20]. According to NEGOPY, we define vertex types as isolate, dyad, liaison, and group member. As we deal with undirected networks, we do not classify tree-nodes. According to Richards, an isolate is an individual with at maximum one friend. Two persons connected only to each other are denoted as dyad. Liaisons are individuals with more than 50% connections to members of different groups. Liaisons can also be nodes that are mostly connected to other liaisons and with less than 50% links to group members.

A composition of minimum three individuals is referred to as group, if the individuals share more than 50% of their linkage, build a connected component and stay connected if up to 10% of the group members are removed. For measuring the quality, the number of positive vertices $n_k^p$ in each friendship-pattern class $k$ is calculated after each iteration of the simulation. Subsequently, the error $e_k$ is calculated as difference between $n_k^p$ of simulated and real values. The $average - error$ $e$ denotes the weighted average error of the simulation and $n_k$ the number of vertices in friendship-pattern $k$ :

$$e = \frac{\sum_{k=1}^{n} e_k * n_k}{\sum_{k=1}^{n} n_k} \tag{10}$$

$(iv)$ The average estimation error $\epsilon$ assesses the average difference between simulated values for behavior and real behavioral outcomes. Here $n$ denotes the total number of nodes in the simulation, while the difference between simulation and reality for node $i$ is represented by $\epsilon_i$.

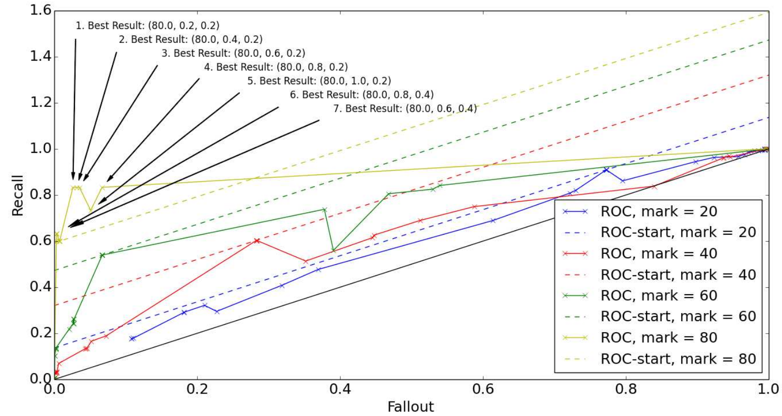
$$\epsilon = \frac{\sum_{i=1}^{n} \epsilon_i}{n} \tag{11}$$

### 4.3 Results

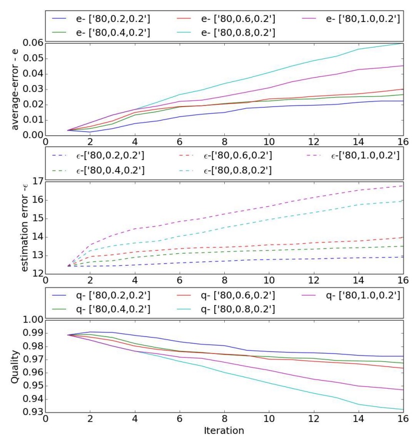
This subsection presents the results from the experiments presented earlier. We first present results for the experiments with FUNDAJ data and subsequently report results for experiments with the Scottish data set.

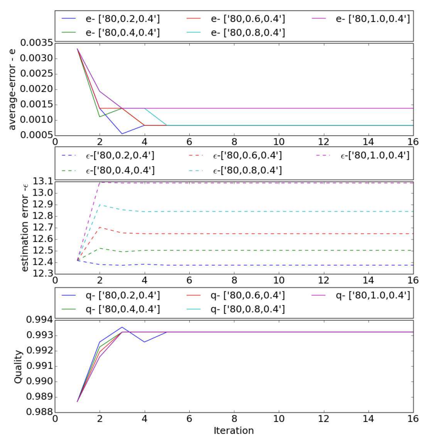
**Results - FUNDAJ** Figure 1 presents the results for simulations with FUNDAJ data for 15 iterations. Figure 1(a) contains ROC-curves for the experimental results with varying settings of $tm$, $GT$, and $f$. For each investigated value of mark threshold $tm$, the Figure illustrates an individual ROC-curve. The dashed lines indicate the ROC-level of the respective setting for $tm$ before starting the simulation. Thus only parameter settings leading to ROC-values situated above the respective dashed line can be considered as settings that improve the quality of the simulation. The colored lines in Fig. 1(b) represent the development of quality indicators $q$ and $e$ for distinct parameter settings and also indicate the average estimation error $\epsilon$ during the run-time of the simulation.

The results with the highest $Youden-Index$ in simulations with FUNDAJ data set are indicated by arrows pointing from the respective parameter settings for mark-threshold $tm$, global threshold $GT$ and adaption parameter $f$ in parentheses as ($tm$,$GT$,$f$) in Fig. 1(a). The results for $q$, $e$ and $\epsilon$ of those most promising parameter settings are presented in Fig 1(b) and 1(c). The more detailed analysis of the five parameter settings that were performing best in ROC-curve analysis in Fig. 1(b) yields increasing $e$ and increasing estimation error $\epsilon$ while $q$ continuously decreases. However, as presented in Fig. 1(c) the second best performing parameter settings from ROC-curve analysis lead in general to a decay of $e$ and significant growth of $q$ whereas at least one setting (80,0.2,0.4) also decreases estimation error $\epsilon$ slightly.




(a) ROC-Curve FUNDAJ




(b) Analysis FUNDAJ-$f = 0.2$ (c) Analysis FUNDAJ $f = 0.4$

Fig. 1: ROC-Curve and Analysis for Coordination Game simulations with FUNDAJ data - 15 iterations; For varying threshold marks $tm$, classifying the pupils as *good students*, if their mark is greater then $tm$ or *bad students* if their performance is below $tm$ and for varying settings of $GT$ and $f$, as pointed out in parentheses ($tm$,$GT$,$f$). $Recall = \frac{true-positives}{n^p}$; $Fallout = \frac{false-positives}{n^n}$.

**Results - Scottish data** Figure 2 illustrates the results for experiments with the Scottish data set for 50 iterations for each of the investigated behaviors. The solid lines in Fig. 2(a) illustrate the $Recall$-$Fallout$ relation for varying parameter settings and for different behaviors. The black diagonal line in this graph indicates $Recall$-$Fallout$ ratios that represent random processes, while the dashed lines indicate the ROC-level of the start situation. Since experiments with Scottish data were run with two different networks as explained in Sect. 3, analysis of $q$, $e$ and $\epsilon$ in Fig. 2(b), 2(c), 2(d) and 2(e) contain blue lines, indicating the values calculated in relation to t+1 and red lines, representing the results calculated in relation to *benchmark t+2*. ROC-curve for the simulation of diffusion of behavior $sport$ in Fig. 2(a) is very close to the diagonal of the graph, indicating that the simulation is rather a random process. Furthermore ROC-values cannot reach the ROC-level before starting the simulation indicated by the dashed line. However, there are two values for $t$ that yield ROC-values above the diagonal of which $t$ = 0.55 generates the most promising results. Hence, $q$, $e$ and $\epsilon$ development are analyzed over the whole run time in Fig. 2(b). It is observable, that $e$ in $t + 1$ indicated by the blue line decreases significantly until the 25th iteration, which is when the network $g_t$ is replaced by network $g_{t+1}$. After the 25th iteration, $e$ in $t + 2$ decreases heavily. $\epsilon$ decreases slightly for benchmark $t + 2$ but increases if compared to benchmark $t + 1$. Although decreasing for the first five iterations, $q$ remains stable during the following 20 iterations and slightly improves after 25 iterations.

ROC-curve for smoking behavior in Fig. 2(a) yields positive results for $t$ 0.35, 0.4 and 0.45, significantly outperforming the initial ROC-value indicated by the dashed line. A deeper examination of $q$,$e$ and $\epsilon$ development during run time in Fig. 2(c) shows that as compared with benchmark $t+1$ neither $q$, nor $e$ or $\epsilon$ develop positively. Though, compared with benchmark $t + 2$ a strong improvement of $q$, as well as a significant decrease of $e$ and a slight decrease of $\epsilon$ is observable.

The $t$ values indicated by the ROC-curve for Alcohol-use in Fig. 2(a) do not reach initial ROC-level and yield decreasing $q$ and increasing $e$ until the underlying network is changed after 25 iterations, initiating a slight improvement of those values for both benchmark values as presented in Fig. 2(d). Nevertheless, $q$ never reaches a value higher than the start value, also $e$ does not drop under its start value and $\epsilon$ remains on an equal level. ROC-curve for Drug-use in Fig. 2(a) yields positive results for $t$ 0.35, 0.4 and 0.45, slightly exceeding the initial ROC-value. Figure 2(e) presents decreasing $e$ and $\epsilon$, as well as increasing $q$ over the run time for benchmark value $t + 2$, while all quality measures develop negatively for benchmark $t + 1$.

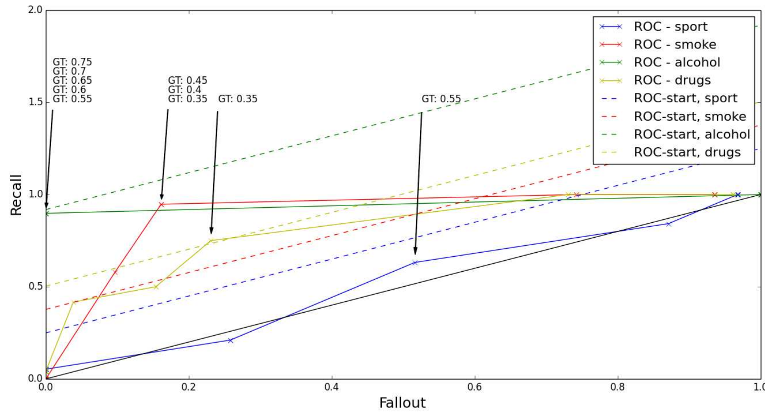
## 5 Discussion

### 5.1 FUNDAJ data set

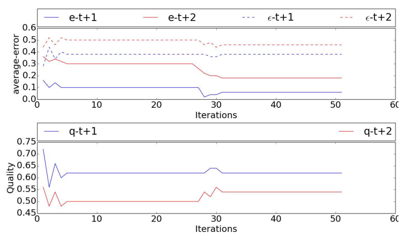
As pointed out in Sect. 4, the parameter setting (80,0.2,0.4) performs best as under this setting average-error $e$ is being more than halved (approximately 75%). For this setting also match quality $q$ increases slightly, $\epsilon$ shows a small decay, and $Youden - Index$ improves. This indicates that the setting reasonably approximates the real system state. However, simulation is not very adequate in estimating the individual behavior. Thus we argue that diffusion of marks can be reasonably modeled as a coordination game if the researcher is willing to disregard individual states and is interested in the global state of the network instead. Results further indicate that 15 iterations under the given parameter setting are good for approximating one school year.

### 5.2 Scottish data set

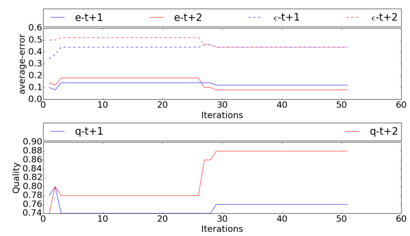
Simulating the coordination-game spread for behavior $sport$ with $GT = 0.55$ yields a relatively small $Youden - Index$ and cannot improve the ROC-level of the initial situation. However, the development of $average - error$ for benchmark t+1 and t+2 yield



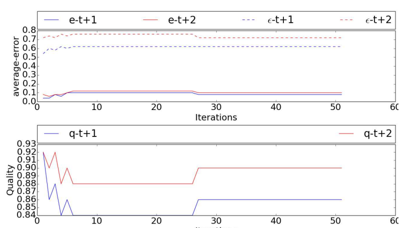

(a) ROC-Curves Scottish Data



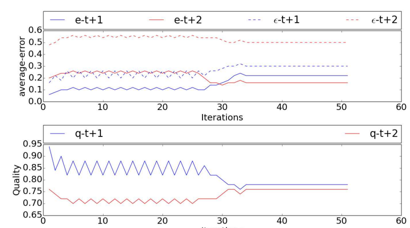

(b) Analysis Sport - $GT = 0.55$




(c) Analysis Smoking - $GT = 0.45$




(d) Analysis Alcohol - $GT = 0.65$




(e) Analysis Drugs - $GT = 0.35$

Fig. 2: ROC-curves and Analysis for Coordination Game simulations with Scottish data set - 50 iterations. For varying behaviors and for varying settings of $GT$. $Recall = \frac{true-positives}{n^p}$; $Fallout = \frac{false-positives}{n^n}$.

improvement of the overall state of the network, while decreasing $match - quality\ q$ and increasing $\epsilon$ for both benchmarks. Although improving the estimation of the general network state, setting $GT = 0.55$ cannot improve the estimation quality and can therefore not be considered a good setting for $GT$.

As for simulating the spread of behavior $smoke$ throughout the given network, we found strong evidence for the suitability of parameter $GT$ = 0.35, 0.4 and 0.45 in the ROC-curve. Run time analysis of $e$, $\epsilon$ and $q$ indicate that the parameter setting when run on network $g_1$ cannot reproduce the spreading during the first year, since the former grows while the latter declines for the first 25 iterations. However considering benchmarks t+2 and network $g_2$, all three quality indicators support the hypothesis that spreading occurs as a coordination game with $t$ = 0.35, 0.4 or 0.45. Recall, that children were around age 13 when the study started, this discrepancy may be explained by the nature of the behavior smoking, which probably has a higher attraction to children aged 14 to 15 than to children aged 12 to 13. Similar but not as striking evidence can be found when examining behavior $drug - use$. As $drug - use$ has been explicitly surveyed as the use of cannabis, this seems coherent, since tobacco use does commonly precede cannabis use. Conversely, for the behavior $alcohol - use$, results are not clear. ROC-curves indicate that parameter settings yielding reasonable estimates of the real situation exist. Yet, run time analysis of those cases show that those promising parameter settings do not lead to an improvement of the estimation. Hence, we argue that for alcohol-use we cannot find evidence that spreading of behavior can be modeled as a coordination game within the given data set. This might also be related to the age of the students, since parents influence might be stronger during this period. Additionally due to the restriction of available data to female students the lack of spreading could be gender related.

## 6 Conclusion

In this paper we adopted the coordination-game mechanism for simulating the spreading process of behavior throughout social networks. We ran the simulation on two different data sets, the FUNDAJ study with school children from metropolitan area of Recife and the study from Scottish female pupils. We investigated the spread of behavior "commitment to school education" represented by the marks of the pupils in the FUNDAJ study, as well as the behaviors "Substance use" for tobacco, drugs and alcohol and the behavior "practicing sports" as surveyed in the Scottish data set. We found good indications that a coordination-game mechanism underlies the spread of behavior "commitment to school education" as well as "smoking" and "drug-use" but could not find comparable evidence for behavior "Alcohol-use". Results for behavior "practicing sports" were not clear. We argue that the missing evidence for behavior "Alcohol-use" may stem from the nature of the data set, since surveyed individuals were below 16 years of age until the end of the survey. Moreover only female pupils participated. Since male adolescents are more susceptible to early alcohol-use, this could be an explanation for the lack of evidence, for that particular aspect.

This work serves as a first step in simulating the spread of behavior throughout social networks, since it provides evidence that (1) there is an underlying game-environment for the agents within the social system (2) that it can be modeled as a coordination game. However, the players of this game, the bounded rational agents [21] might be equipped with decision finding mechanisms that better approximate human decision making. Though driving the social systems from a real start situation towards the state in reality after one or respectively two years, the investigated deterministic mechanism still lead to a considerable difference between the real and the simulated system. Hence, we argue that the deterministic mechanism is not fully capable to simulate human bounded rationality and the lack of information humans face within their decision process. Besides this, eventual noise within the data and external influences may not be represented by a deterministic mechanism. Future work should therefore deal with the creation of a heuristic decision mechanism for the individual agents, that better represents human decision making within a coordination-game setting. In addition, a binary behavioral variable is an extreme simplification for the on continuous scales measured nuances of human behavior such as sports activities, drug- and alcohol consumption or school performance. It is therefore desirable to investigate how more complex scaling systems influence the outcomes of this research. Furthermore, inter-temporal components shall be introduced, representing an "aging" of relations and behaviors, modifying the influence of neighbors according to the "age" of the friendship, as well as according to the past behavior of the neighbor. In the same sense, friendship weights may be modified according to the position of friends within the individual networks since people may tend to follow "role models". Finally, the presented mechanism must be applied to different data sets in order to empirically verify the results.

**Acknowledgements.** This research was partially funded by Andrea von Braun Stiftung. The authors thank Fundação Joaquim Nabuco for providing data and domain expertise.

## References

1. N. A. Christakis and J. H. Fowler, *Connected: The surprising power of our social networks and how they shape our lives*. Little, Brown, 2009.
2. D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.
3. M. Lelarge, "Diffusion and cascading behavior in random networks," *Games and Economic Behavior*, vol. 75, no. 2, pp. 752–775, 2012.
4. B. Latane, "The psychology of social impact." *American psychologist*, vol. 36, no. 4, p. 343, 1981.
5. N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *New England journal of medicine*, vol. 357, no. 4, pp. 370–379, 2007.
6. J. J. Jordan, D. G. Rand, S. Arbesman, J. H. Fowler, and N. A. Christakis, "Contagion of cooperation in static and fluid social networks," *PloS one*, vol. 8, no. 6, p. e66199, 2013.
7. A. D. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. 24, pp. 8788–8790, 2014.
8. T. A. Brakefield, S. C. Mednick, H. W. Wilson, J.-E. De Neve, N. A. Christakis, and J. H. Fowler, "Same-sex sexual attraction does not spread in adolescent social networks," *Archives of sexual behavior*, vol. 43, no. 2, pp. 335–344, 2014.
9. E. Marques, *Redes sociais, segregação e pobreza*. Editora Unesp, 2010.
10. J. Kratzer and C. Lettl, "Distinctive roles of lead users and opinion leaders in the social networks of schoolchildren," *Journal of Consumer Research*, vol. 36, no. 4, pp. 646–659, 2009.
11. E. L. Paluck and H. Shepherd, "The salience of social referents: A field experiment on collective norms and harassment behavior in a school social network." *Journal of personality and social psychology*, vol. 103, no. 6, p. 899, 2012.
12. N. E. Friedkin, *A structural theory of social influence*. Cambridge University Press, 2006, vol. 13.
13. M. Granovetter, "Threshold models of collective behavior," *American journal of sociology*, pp. 1420–1443, 1978.

14. M. Granovetter and R. Soong, "Threshold models of interpersonal effects in consumer demand," *Journal of economic behavior & organization*, vol. 7, no. 1, pp. 83–99, 1986.
15. T. C. Schelling, "Dynamic models of segregation," *Journal of mathematical sociology*, vol. 1, no. 2, pp. 143–186, 1971.
16. Coordenação de Estudos Econmicos e Populacionais, Fundação Joaquim Nabuco  Fundaj, "Determinantes do desempenho escolar na rede de ensino fundamental do Recife," 2013.
17. L. Michell and A. Amos, "Teenage friends and lifestyle study dataset," 1997.
18. T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.
19. W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.
20. W. D. Richards and R. E. Rice, "The negopy network analysis program," *Social Networks*, vol. 3, no. 3, pp. 215–223, 1981.
21. P. M. Todd and G. Gigerenzer, "Bounding rationality to the world," *Journal of Economic Psychology*, vol. 24, no. 2, pp. 143–165, 2003.