Working paper GAEL n° 02/2022

Consumption - Energy - Innovation

Spreading active transportation: peer effects and key players in the workplace

Mathieu Lambotte
Sandrine Mathy
Anna Risch
Carole Treibich

October 25, 2022

JEL: D91, R41, C31

https://gael.univ-grenoble-alpes.fr/accueil-gael

Spreading active transportation: peer effects and key players in the workplace

Mathieu Lambotte[1,*], Sandrine Mathy[1], Anna Risch[1], and Carole Treibich[1]

[1]Univ. Grenoble Alpes, CNRS, INRAE, Grenoble INP, GAEL, 38000 Grenoble, France

[*]Corresponding author: Mathieu Lambotte, mathieu.lambotte@univ-grenobles-alpes.fr, 04 76 74 29 27, bureau 201, BATEG, 1241 rue des Résidences, 38400 Saint-Martin-d'Hères

July 13, 2022

Abstract

We investigate the role of peer effects at the work place on the individual choice of transportation mode. We collect original data through an online survey on networks and sustainable behaviors among 334 individuals working in ten laboratories of the University of Grenoble Alps in February 2020. Using a linear-in-means model for binary outcomes and distinguishing endogenous and exogenous peer effects, correlated effects and network endogeneity, we find that peers have a significant and positive effect on individual active transportation mode's choice. We show that in our setting, a simulated policy or intervention would be almost twice more effective in spreading active transportation mode through social spillover effects if it targets *key players* rather than random individuals.

Keywords— peer effects, social network, workplace, transportation choice, key players

JEL Classification— D91, R41, C31

Acknowledgement

We would like to thank Paolo Crosetto, Habiba Djebbari, Aristide Houndetoungan and Yann Bramoullé for their comments at different stages of this article. We also thank the participants of the International Workshop of Spatial Econometrics and Statistics, of the JMA, the Congrès of the AFSE and the Journées Louis-André Gérard-Varet. This work was partially funded by IRS IDEX UGA RESMOB and CDP IDEX UGA MOBILAIR (ANR-15-IDEX-02). All authors declare that they have no conflicts of interest.

1 Introduction

Reducing the use of private cars is a priority because of health and environmental concerns. Indeed, road traffic is one of the main contributors to greenhouse gas emissions, reinforcing climate change. Emissions of atmospheric pollutants have serious health consequences in terms of mortality and morbidity. An important use of private cars also leads to noise and congestion issues as well as a sedentary lifestyle, which has become an important public health issue, contributing to obesity and lack of physical activity. Thus, inducing modal shifts to active transportation modes would reduce the negative environmental and economics externalities of road traffic (pollution, road wearing or accidents (Bouscasse et al., 2022)) and the prevalence of several diseases linked to inactivity (cardio-vascular diseases, Type 2 diabetes, colon-rectal and breast cancers (Rojas-Rueda et al., 2013)).

While several surveys show that many people are aware of the benefits of reducing car use for environmental or health reasons (IPSOS and BCG, 2017), in many cases intentions do not translate into actual behavioral changes (Sheeran and Webb, 2016). In 2017, 74% of French working people who commute to work used their car, 16% took public transport and only 8% walked or biked (Brutel and Pages, 2021). Even for distances of less than five kilometres, which might easily be traveled by walk or bicycle, car travels still account for 60% of home-work commuting trips.

Public authorities are trying to reduce the use of cars by introducing various measures to encourage the uptake of alternative modes. For example, since 2010 in France, employers must reimburse 50% of public transport annual passes. Since 2021, a so-called *sustainable mobility package* ("forfait mobilité durable") enables companies to voluntarily and monetary compensate their employees for travelling to work by other means than car. These two schemes are cumulative and exempted of taxes up to 500 euros per year per employee. Despite the existence of these legal schemes and the development of bicycle and public transport infrastructures, uptake of these schemes and resulting transportation behavior changes are low.

Indeed, behavioural changes in transport modes are particularly difficult to initiate. Individual behavior depends on many intertwined determinants that may be economic, social or psychological as well as constrained by the facilities, institution and infrastructures that surround us. Nevertheless, social norms can be effective in disseminating pro-social behaviors (e.g., the use of seat belts). Beyond that, our social environment can also play a significant role in explaining individual behaviors and associated changes. According to Akerlof (1997), economic analysis ought to shift its focus from methodological individualism to broaden the concept of utility maximisation, accounting for the social distance to the social groups' norm, i.e. accounting for social pressure or conformity desire. For example, Fortin and Yazbeck (2015) have shown the role of conformity and friendly relationships of adolescents in the frequentation of fast-food restaurants and thus in the obesity epidemic. Powell et al. (2005) have similarly studied conformity in smoking behavior among adolescents. These peer effects are important to analyse because they can promote the dissemination of good practices (changing the social norm, i.e. in our case the average behavior of a social group). Peer effects are then capable of amplifying the impact of public policies beyond the direct effect of policies or interventions, thus acting as a social multiplier.

While most people spend a large part of their daily lives at their workplace, the diffusion of sustainable behaviors in this key arena of social interactions has not been extensively studied. Furthermore, social links with coworkers are not perfect substitutes to links with family or neighbors (Videras et al., 2011) and one might also want to explore the impacts of coworkers on personal behaviors. To the best of our knowledge, peer effects in the workplace have almost exclusively aimed at evaluating the impact of colleagues' productivity on individual's productivity (Lindquist et al., 2015). However, given the time spent in the workplace, regular and friendly exchanges with colleagues during coffee breaks and lunches taken together may lead to the adoption of new practices in non-work related areas. In the context of home-to-work mobility, for example, it is possible that following discussions with colleagues, some employees will be encouraged to use a mode of transportation other than the car. This social influence can be driven either by conformity (if a large number of colleagues are using a specific transport mode), by exchanges of information on carpooling opportunities, in-house company policies or on fast and pleasant routes (i.e. strategic complementarity). Companies or administrations thus appear to be potentially favourable places in which to focus public policy efforts so that more environmentally friendly mobility behaviors emerge and then spread to the private sphere.

The analysis of peer effects on the active transportation mode at the workplace proposed here aims at assessing the extent to which interactions with colleagues can lead to the adoption of an active transportation mode. We deployed a survey among ten research laboratories (henceforth labs) on the university campus of Grenoble Alps that allowed the construction of an original database gathering information on individual behaviors and their professional network. In order to estimate peers effect, we follow the linear-in-means model à la Bramoullé et al. (2009). This model estimates social influence as the mean behavior of peers, i.e. an estimate of the social norm in a given social group. We show that conformity has a strong influence in active transportation mode's choice and that an individual whose peers all use an active transportation mode has a 50 to 70 percentage points higher probability of using herself an active mode compared to an individual for which none of her peers use an active transportation choice. Using the estimated coefficients from the linear-in-means model, we predict, for individuals that live relatively close from work and do not already commute using an active transportation mode, the aggregated increase in the probability of using an active transportation mode following a public policy or intervention making 10% of the interviewees shift from a non-active to an active transportation mode. If those 10% are randomly selected, endogenous peer effects create a *spillover* effect, increasing by 5.8% the probability that non-recipients use an active transportation mode. We show that if the public policy or intervention targets well-connected and influent individuals (i.e. *key players*), based on their in-degree centrality or the influence they have on their peers, instead of randomly selecting the recipients, the spillover effect on non-recipients almost doubles.

The rest of the paper is organised as follows. Section 2 is a literature review on transportation modes' choice and peer effects. Section 3 presents the study protocol, the data collection process and descriptive statistics. Section 4 displays the empirical and identification strategies. Section 5 reports the results of the estimations of peer effects and presents robustness analyses. Section 6 displays a key player analysis of the effect of an intervention leading to shifts toward active transportation modes. Finally, section 7 discusses the results, the methodological challenges of the study, its limitations and future research opportunities.

2 Literature review

2.1 Literature on social influence in transportation mode choice

Modal choice is a complex decision process determined by a wide variety of factors stemming from different dimensions. Indeed, transportation mode is influenced by the cultural and social contexts (social norm and conformism), individual characteristics (age, gender, physical capabilities, cycling experience for instance) as well as the specifics of the trip (speed, cost, availability of alternatives, congestion, etc) (Sherwin et al., 2014). De Witte et al. (2013) review 76 articles on transportation modes' choice and classify its determinants into four groups: the determinants frequently studied and usually found significant (such as car availability, income, density, household type), the ones rarely studied and frequently found significant (such as habits, trip chaining, parking), those often studied and rarely found significant (such as gender, employment, motive) and finally the ones rarely studied and infrequently found significant (such as interchange, information, perceptions).

This classification nevertheless omits social influence such as peer effects, which are rarely studied in transportation modes' choice but are usually found significant in other topics. Indeed, the influence of social relationships has been intensively studied on different behaviors and outcomes in daily life, among which academic achievement (Boucher et al., 2014; Calvo-Armengol et al., 2009; Landini et al., 2016), eating behaviors and obesity (Dieye and Fortin, 2017; Fortin and Yazbeck, 2015) and productivity at work (Beugnot et al., 2019; Lindquist et al., 2015).

Accordingly, transportation research has recently begun a shift from an individual focus to a social focus, and behavioral models have started to account for social context. In particular, Maness et al. (2015) propose a literature review of social influence in transportation research. They highlight that two forms of social influence mechanisms have been analyzed in travel behavior models: conformity - an endogenous social influence mechanism - and compliance - a contextual social influence mechanism.

Conformity is the primary social influence mechanism modeled in transportation. An individual is said to conform when she adapts its own behavior to that of the majority of its peers, i.e. to the social norm, in particular "to develop and preserve meaningful social relationships, and to maintain a favorable self-concept" (Cialdini and Goldstein, 2004, p. 591). Experiments have shown that individuals widely shift from car to bicycle when commuting with a bicycle becomes the social norm (Kormos et al., 2015). Similarly, a qualitative research in England showed that most of the interviewees that were new bicyclists were influenced by their peers to change their transportation mode (Sherwin et al., 2014). In contrast with conformity, compliance describes the behavioral changes stemming from advice or commands from peers. These triggers can be explicit (e.g. direct request from a supervisor) or implicit (e.g. an advertisement).

Pike and Lubell (2016) explore the effects of social influence on biking and apply an instrumental variables approach to address the potential endogeneity related to similarities in the choice environments of socially connected individuals as trip characteristics or transit access may influence multiple members of a social network at the same time. They conclude that social influence is strong in bicycle transportation mode, even if one accounts for the correlation in behaviors due to a common environment. This is in line with the findings of other research that has also addressed spatial sources of endogeneity in social effects related to several transportation modes (Dugundji and Gulyás, 2008; Goetzke and Andrade, 2010; Goetzke and Weinberger, 2012; Kim et al., 2017; Phithakkitnukoon et al., 2017; Walker et al., 2005; Wang et al., 2015).

Pike and Lubell (2018) and Phithakkitnukoon et al. (2017) refine this finding as they state that the effects of social influence are nuanced by contextual factors, such as commuting distance: the strength of social influence is lower for those with longer commuting distances as biking is more costly (utility-wise) than driving. Social influence is most important when the external trip characteristics entail relatively equal travel costs for different modes (Pike and Lubell, 2018).

We find two main limits to the previous literature on social influence in transportation modes. Firstly, most studies of social influence in transportation modes' choice analyse neighbourhood networks (networks where individuals are assumed to be connected if they live near one another, share socio-economic traits, or may otherwise be grouped by a relevant characteristic) instead of the usual peers networks (family, friends, coworkers). However, the influence from peer networks and neighbourhood networks operates in different ways. According to Pike (2014) identifiable peers' connections are more relevant in the decision to commute via bicycle, while neighbourhood networks are more relevant in the decision to take public transports. We collect peer network data in order to refine these results. Secondly, the rich methodology literature steaming from networks econometrics is underused in transportation choices' models. Especially, the different sources of social influence are rarely separated and the marginal effects of peer effects on individual transportation choices are scarcely estimated. However, using peer networks data and a linear-in-means model of peer effects may estimate such causal peer effects.

2.2 Relevance of the network econometrics approach

Recently, the economic literature has integrated new perspectives into the empirical study of individual behavior by considering the effects of peers' choices or behaviors. Family, friends or colleagues can influence individual decisions, for example through conformity desire and mimicry. These social interactions are likely to play an important role in the effectiveness of public policies and intervention, by amplifying their effects (snowball effect, contagion or social propagation), or on the contrary by attenuating them. For example, an increase in the price of cigarettes has a direct effect on reducing a smoker's tobacco consumption via the classic "price" effect. It can also have an indirect effect through the decisions of her entourage (friends, family and coworkers) who may also decrease their consumption, thereby reducing the opportunities to smoke (see Powell et al. (2005) who studied the impact of friendship network among adolescents' smoking behavior). Similarly, researchers interested in overweight have highlighted the impact of the friendship network on the consumption of fast food among adolescents. They show that an increase in friends' average consumption of fast food encourages adolescents to increase their own fast food consumption (Fortin and Yazbeck, 2015). Through the estimation of a social multiplier, which measures the contagion effect of an exogenous shock, the authors show that peer effects amplify the impact of policies aimed at combating overweight problems. We apply here the linear-in-means model of Bramoullé et al. (2009) to estimate interpretable marginal effects in the presence of a binary dependent variable (Boucher and Bramoullé, 2020; Fletcher, 2012).

2.3 Identification issues

Social influence and peer effects remain empirically understudied for a lot of behaviors, due to the lack of data on social networks or on the behaviors adopted by network members. In addition, the empirical study of the effects of social networks on individual behavior raises several econometric issues. According to Manski (1993) and the extensive literature that followed, social influence on individual behavior can be disentangled in three distinct effects: endogenous effects, exogenous (contextual) effects and correlated effects.

Endogenous peer effects are the most straightforward source of social influence on one's behavior. Indeed, these peer effects incorporate conformity desire, the fact that individuals tend to take into account their peers' behaviors. Network econometrics models generally considered that the social norm individuals face is the average behavior of their peers.

Contextual effects, also referred to as exogenous peer effects or compliance, capture the fact that peers' characteristics may influence one's behavior, i.e. biking is well-perceived among high income peers, and so peers' income would influence one's transportation mode if she cares about how she is perceived by her peers.

Correlated effects can be understood as the propensity of peers to behave similarly because they either share common individual characteristics or face a common environment (Bramoullé et al., 2009; Manski, 1993). Indeed, in most cases, we can assume that individuals select the people with whom they interact and therefore share common characteristics or interests, as homophily behaviors are dominant in peers' selection and network formation. In addition, individuals in a given workplace share a same environment as they work at the same place and thus face the same set of incentives or external shocks. However, some of these individuals characteristics or common environment variables may be unobserved to the econometrician, resulting in an omitted variable bias of the peer effects. Indeed, if these unobserved variables are not accounted for, some of their effects might be confounded with the peer effects we are interested in. Considering this common environment and network endogeneity, we can expect that individuals in a same network will act in a similar way and therefore are likely to have positively correlated outcomes even in the absence of true peer effects .

Investigating the impact of social relationships therefore implies the methodological challenge of distinguishing the impacts of social interactions among these three effects: endogenous and exogenous peer effects and correlated effects.

In addition, Manski (1993) has highlighted the issues related to estimating causal peer effects because of the so-called "reflection problem". Indeed, it is difficult to separate the effect of one's entourage on its behavior and the effect of the individual's behavior on her peers. This simultaneity causes an endogeneity problem in the estimation. To overcome this issue, Bramoullé et al. (2009, 2019) show that the characteristics of the peers of peers which are not directly peers (i.e. a given individual, its peers, and the peers of its peers form an intransitive triad) can be valid instruments for one's peers behaviors. Indeed, peers of peers characteristics are strongly correlated with peers' behavior but are uncorrelated with the behavior of this given individual. They propose an approach to address these identification problems and to estimate the peer effects in so-called *linear-in-means* models.

3 Data collection and descriptive statistics

3.1 Census and social network survey

In this research we aim at studying the impact of the work environment on transportation choice. More precisely, we wonder whether colleagues from the workplace influence behaviors that are not directly related to work or the workplace but to personal choice. To carry out this study, we set up an original survey on the university campus of Grenoble Alps. Data collection is described below.

One of the challenges we face is the construction of the networks per se. While the Grenoble University hosts more than 90 research structures on two distinct geographical sites, we drastically restricted the number of research labs in which we wanted to deploy the survey in order to maximise the share of each structure's members participating to the study. Obtaining a high response rate per laboratory is a crucial objective of the data collection phase, in order to be able to trace the social interactions within each research unit, and properly represent the professional network of each respondent (i.e. limiting the partiality of the professional network). The higher the response rate, the better we will be able to minimize bias in the analysis.

Therefore, the selection of laboratories was based on two criteria: (i) the laboratory must be located on the main university campus and have only one geographical site and (ii) very large laboratories are excluded. This resulted in a selection of 14 laboratories. In a second step, the director of each selected research structure was contacted by email and/or phone to present him/her the study, confirm that the structure fulfilled the two inclusion criteria and ask whether he/she would agree to us coming to present the study in the laboratory to all the staff personnel. This second phase resulted in a final selection of 10 laboratories varying in size from 29 to about 300 persons (including permanent and non-permanent staff, research and research support staff, excluding non-doctoral students). All in all, this represents 1,335 workers. The organised information meeting about the study in each laboratory was the occasion to distribute flyers presenting the study. This visit was followed by an email sent to each member of the selected research structures, with the assistance of the direction, recalling the purpose of the study and providing the consent form to be filled in order to receive the baseline questionnaire. The collection of consent forms started on January 7, 2020 and ended on January 27, 2020. Following the *General Data Protection Regulation* guidelines, only staff who provided their consent and thus agreed to participate in the survey received an email containing a link to the "baseline" questionnaire[1]. This survey has been carried out using the Sphinx software, the questionnaire is therefore completed online by each respondent. This survey aims at collecting information on the following elements:

- **The networks.** We ask participants to name the people with whom they talk the most and regularly in their workplace. A list of colleagues appeared in the questionnaire. Only individuals who have consented to participate to the study appear in this list. The goal is to identify each individual's network to see if their peers have an effect on their behavior.
- **Their behavior.** The objective is to see if the social interactions at the workplace have an effect on individual's behaviors concerning transportation choice.
- **Their socio-demographic characteristics** in order to control for individual heterogeneity.

The baseline questionnaire was sent on February 6, 2020 to the 465 individuals who agreed to participate to the study. A first reminder was sent on February 11, 2020 and a second one on February 18, 2020. 407 individuals actually completed the baseline questionnaire. Out of these 407 individuals, some did not report social network information or are isolated (i.e. do not have any peers among the other participants), yielding 334 exploitable observations for the estimations. Table 1 reports the consent rate and the participation rate for each research structure.

[1] Available at https://osf.io/nqat4/

The protocol has been approved by the multidisciplinary ethics committee of the University of Grenoble Alps) and complies with the *General Data Protection Regulation*. All the document related to the survey were presented in both French and English.

Table 1: Survey

Labs (acronym)	3SR	CRJ	IAB	IGE	LEG	LEP	LGP	LIP	SENS	SIM	Total
Nb of people in the lab †	110	77	300	227	100	121	77	182	29	112	1,335
Nb of people consenting to participate	54	38	41	105	33	46	38	44	13	53	465
Nb of people fulfilling baseline	51	28	34	94	30	39	34	40	12	45	407
Nb of surveys fully exploitable	47	20	22	86	27	28	22	39	10	33	334
Potential participation rate ‡	46%	36%	11%	41%	33%	32%	44%	22%	41%	40%	35%
Actual participation rate ∓	43%	26%	7%	38%	27%	23%	29%	21%	35%	30%	25%

Notes: † the number of people in the research structures is an estimation of at the time of the survey. ‡ The potential participation rate corresponds to the share of individuals consenting to participate who actually fulfilled the questionnaire.∓ The actual participation rate corresponds to the share of individuals who correctly reported network information and are not isolated, thus considered in the analysis.

Our database contains ten different professional networks composed of 12 to 94 nodes (individuals, Table 2). While individuals could cite up to six colleagues, the average number of peers varies from 2.14 to 4 depending on the research structure. We face a problem of completeness of the networks as only labs' members who agreed to participate to the study appear on the list of colleagues that could be cited (cf. Table 1). Therefore, many respondents declare that they would like to cite a colleague that does not participate to the study.

Table 2: Network characteristics

Labs (acronym)	3SR	CRJ	IAB	IGE	LEG	LEP	LGP	LIP	SENS	SIM
Nodes∓	51	28	34	94	30	39	34	40	12	45
Edges±	175	60	100	373	120	105	117	142	31	120
Number of individuals cited on average by a colleague∗	3.43	2.14	2.94	3.97	4.00	2.69	3.44	3.55	2.58	2.67
Share of people wishing to name someone who did not participate in the survey	82%	59%	88%	70%	97%	64%	74%	93%	75%	82%
Potentials links	2550	756	1122	8742	870	1482	1122	1560	132	1980
Density‡	0.069	0.079	0.089	0.043	0.138	0.071	0.104	0.091	0.235	0.061
Reciprocity†	0.336	0.273	0.44	0.343	0.383	0.438	0.297	0.404	0.737	0.278
Transitivity∓	0.273	0.657	0.809	0.424	0.338	0.279	0.587	0.522	0.260	0.684

Notes: ∓ Number of individuals, ± Number of links, ∗ edges/Nodes
‡ The density is equal to the number of existing links/number of potential links.
† The reciprocity is equal to the number of mutual links/number of existing links.
∓ The transitivity is equal to the number of triangle in the graph/number of connected triple of nodes.

Given that some respondents did not correctly fill the socio-demographics entries or networks' information, the final data used for the empirical analysis reaches 334 individuals. The networks formed by these 334 individuals is represented in Figure A.1 in the Appendix.

3.2 Descriptive statistics

In the survey, we collected socio-economic characteristics: gender, age, number of dependent children (less than 12 years old) in the household, household's income range as well as the urban neighborhood or rural city in which they live. Using the *OpenTripPlanner* software (Malcolm Morgan et al., 2019), we estimate the duration and the distance of the commuting trips (between the lab in which they most often work and the centroid of their neighborhood), for each mode (car, walk, bicycle, public transportation). In addition to the distance from home to work (measured as the commuting distance in car), we create a variable measuring the relative opportunity cost (in time) of using a bicycle to commute instead of a car. This variable, named *Relative opportunity cost of active mode* averages at 27.6%, which means that commuting with a bicycle takes on average 27.6% longer than with a car. Similarly, we compute the relative opportunity cost of using public transportation instead of a car (*Relative opportunity cost of sustainable mode*). We also gather information on health issues that may prevent the use of a bicycle and an index measuring individual's environmental sensibility, based on several questions, used as proxy for pro-environmental behavior. There is 41% of female in the sample with an average age of 40 years. 63% of the sample holds a PhD. The individuals in our sample have on average of 0.52 dependent child and have worked in the research laboratory for 9.6 years (see Table 3). Note that there are no significant differences between the share of individuals that have a driver licence or own a car between the individuals that have an active transportation mode and the others.

Table 3: Descriptive statistics

	Obs	Mean	Std. Dev.	Min	Max
Main transportation mode in the past 3 months					
Car	334	0.24			
Active mode †	334	0.43			
Public transportation	334	0.33			
Sustainable mode ††	334	0.76			
Socio-demographic characteristics					
Male	334	0.59	0.49	0	1
Age in years	334	39.8	11.3	23	75
Years working in the research lab	334	9.5	9.2	0	50
Number of dependent children	334	0.52	0.88	0	3
Holds a PhD	334	0.63	0.48	0	1
Health issue	334	0.05	0.21	0	1
Environmental sensibility index‡	334	0.78	0.17	0	1
Household's monthly revenue <2000€	334	0.22			
Household's monthly revenue 2000 - 4000€	334	0.32			
Household's monthly revenue 4000 - 6000€	334	0.32			
Household's monthly revenue >6000€	334	0.14			
Commuting trips' characteristics					
Commuting distance via car (km)	334	14.9	22.6	0.9	233
Commuting time via car (min)	334	26.2	15.3	5.2	141
Relative opportunity cost of active mode	334	27.6%	92.2%	-56%	429%
Relative opportunity cost of sustainable mode	334	25.9%	89.6%	-69.5%	428.6%

Notes: † the share of people whose transportation mode is bicycle or walk.
†† The share of people whose transportation mode is bicycle, walk or public transportation.
‡ The index is the average of the score to 3 ordered questions:
"How often do you recycle items that you do not need anymore?"
"How often do you sort your household waste?"
"How often do you purchase items with low packaging?"

4 Empirical strategy

In order to measure the influence of peer effects on active transportation mode's choice, we adopt a step-by-step strategy in which we estimate successive models, from the simplest specification to the most complete one. We first model the transportation choice as depending on spatial variables (distance from home to work), socio-demographics characteristics of the individuals (age, gender, number of young children, highest diploma, household's monthly income) and the transportation mode's choice of their peers (endogenous peer effects). In a second model, we add the socio-demographic and spatial variables of one's peers as dependent variables in her own transportation choice. Doing so, we distinguish endogenous (peers' transportation choice) and exogenous (peers' characteristics) peer effects. In a third model, we add fixed effects for each lab, thus accounting for correlated effects, i.e. the common environment faced by individuals from the same lab. In a fourth model, we account for network endogeneity, the fact that link formations among the network might be correlated with the errors terms because of unobserved variables and thus bias the estimation of peer effects (Breza et al., 2020; Houndetoungan, 2020; Hsieh et al., 2020). This final model thus allows us to disentangle endogenous and exogenous peer effects and correlated effects.

More precisely, we adapt the linear-in-means model of peer effects with binary outcomes (Boucher and Bramoullé, 2020) to transportation mode. This model is chosen because it is identified as long as the networks' diameters are greater or equal to 3, i.e. there are individuals that are only indirectly connected, trough a peer of peer. Moreover, this model is easy to implement as well as estimate and can account for correlated effects (labs' fixed effects and network endogeneity in our case). We are interested in active transportation mode's choice, thus the dependent variable is Y_i, a binary variable which take the value of 1 if the individual i uses an active transportation mode for her daily commuting (bicycle or walk) and 0 otherwise (thus if the individual used public transportation or any other motorized transportation mode). The basic linear-in-means model of peer effects can be written in a structural form as:

$$Y_i = \alpha + X_i\gamma + \beta \sum_{j \neq i} g_{ij} Y_j + \epsilon_i, \; i, j = 1, \dots, n. \tag{1}$$

Thus, the active transportation choice of an individual depends on the X matrix of socio-demographics and spatial variables, measured by the vector γ but also on the average of her peers' transportation choices $\sum_j g_{ij}Y_j$, captured by β. a_{ij} is a cell of the adjacency matrix A which represents the network, $a_{ij} = 1$ if individuals i and j are peers, and 0 otherwise. Noting n the number of nodes in the network, i.e. the sample size, A is thus a $n \times n$ matrix and $a_{ii} = 0$, i.e. self-influence is not allowed in the model. Moreover, a_{ij} may be different from a_{ji} as the network is directed: i may consider j a peer whereas j may not. The sum of each i^{th} row in G gives the number of peers of individual i, its outdegree. Each row of A is normalized so that its sum equals unity, which yields the social interaction matrix G with $\sum_j g_{ij} = 1$. $\sum_j g_{ij}Y_j$ is the average of i^{th} peers' transportation choices, i.e. a measure of the social norm in her social group. ϵ_i, the errors, are discrete as Y_i are binary. A key advantage of the linear-in-means model is that the average peers' transportation choice, $\sum_j g_{ij}Y_j$ or for all i, GY, can be instrumentalised by second-degree peers' socio-demographic and spatial variables, G^2X, yielding a reduced form of model 1 (Boucher and Bramoullé, 2020; Bramoullé et al., 2009; Kelejian and Prucha, 1998). These second-degree friends are the peers of i's peers which are not her peers. Without an instrumentation of peers' transportation choices, peer effects' estimate would likely be biased, due to the "reflection problem" (Manski, 1993), the simultaneity between the effect of an individual on her peers' choices and the peers' effects on her own transportation choice. Using such instruments, we proceed with two-stage least squares estimation (2SLS) of all proposed linear-in-means estimators. For brevity, we only present the 2SLS estimation of model 1. Using matrix notation, model 1 can be written as:

$$Y = \alpha + X\gamma + \beta GY + \epsilon$$

With $Z = [1, X, G^2X]$ and *1* a $n \times$ *1* vector of 1, the first step of the classical 2SLS results in the estimation of $GY = Z\pi + \nu$. The 2SLS estimator is obtained using the instrumented $\widehat{GY}$

in the second step:

$$Y = \alpha + X\gamma + \beta\widehat{GY} + \epsilon.$$

In a second model, we distinguish the estimated peer effects between endogenous and exogenous effects by introducing the mean socio-demographic and spatial characteristics of peers in the estimation of one's transportation choice:

$$Y_i = \alpha + X_i\gamma + \beta\sum_{j\neq i} g_{ij}Y_j + \theta\sum_{j\neq i} g_{ij}X_j + \epsilon_i,\ i,j = 1,\ldots,n. \tag{2}$$

In a third model, we introduce labs' fixed effects to control for potential correlated effects, i.e. unobserved variables at the labs' level. Indeed, individuals working in a same lab are likely to face a similar environment that influence their transportation mode in a correlated manner, such as bicycle parking lots:

$$Y_{il} = \alpha_l + X_{il}\gamma + \beta\sum_{j\neq i} g_{ij}Y_{jl} + \theta\sum_{j\neq i} g_{ij}X_{jl} + \epsilon_{il},\ i,j = 1,\ldots,n,\ l = 1,\ldots,10. \tag{3}$$

α_l here are the fixed effects associated to any of the ten labs in our study. In the estimation, we apply a withing-group deviation to Y_{il}, where we subtract the lab average transportation choice instead of including lab dummies to limit the incidental parameter problem (Boucher and Bramoullé, 2020; Bramoullé et al., 2009).

However, peer effects' estimate may still be biased if the network formation is endogenous. Indeed, some unobserved and heterogeneous characteristics, such as gregariousness, may influence both the likelihood to form a link with a colleague and the choice of transportation. More precisely, gregariousness, as an unobserved characteristic, is captured in the error term ϵ_i but also correlated with the network. Such unobserved characteristics thus make the network and the errors correlated. Building on Houndetoungan (2020) and Graham (2017), we use a dyadic link formation model which estimates the probability that two individuals in a given lab are linked, i.e. peers, based on their social distance and their respective gregariousness. Let g^*_{ij} a latent variable such that

$$g^*_{ij} = \Delta X'_{ij}\bar{\gamma} + \mu_i + \mu_j + \epsilon^*_{ij} \tag{4}$$

where $\Delta X'_{ij}$ is a matrix of socio-demographic differences between individuals i and j, μ_i and μ_j are individual effects that capture the degree heterogeneity, i.e. the gregariousness heterogeneity between individuals and ϵ^*_{ij} are the associated errors terms, following a logistic distribution (Houndetoungan, 2020). Assuming that $g_{ij} = 1$ if $g^*_{ij} > 0$, g^*_{ij} thus represents the utility from the link formation. We can write the probability P_{ij} of a link formation between i and j using the logistic distribution:

$$P_{ij} = \frac{exp(\Delta X'_{ij}\bar{\gamma} + \mu_i + \mu_j)}{(1 + exp(\Delta X'_{ij}\bar{\gamma} + \mu_i + \mu_j))} \tag{5}$$

Note that $P_{ij} = P_{ji}$ and thus the link formation model is symmetric. However, the simulated network can still be directed as ϵ^*_{ij} and ϵ^*_{ji} may differ. Fixed effects for each network, i.e. labs, are included in the estimation. Let $\eta_i = (\epsilon_i, \mu_i)$ be a bi-variate normally distributed with $\Sigma_{\mu,\epsilon} = \begin{pmatrix} \sigma^2_\epsilon & \rho\sigma_\epsilon\sigma_{\mu_l} \\ \rho\sigma_{\mu_l}\sigma_\epsilon & \sigma^2_{\mu_l} \end{pmatrix}$ where ρ is the partial correlation between μ_i and ϵ_i. The errors ϵ_i can be rewritten as $\epsilon_i = \rho\sigma_\epsilon\frac{\mu_i - u_{\mu l}}{\sigma_{\mu l}} + \upsilon_i$, where $\upsilon_i \sim \mathcal{N}(0, (1-\rho^2)\sigma^2_\epsilon)$ and $Cov(\mu_i, \upsilon_i) = 0$. The parameters μ_i, which captures individual gregariousness heterogeneity, are estimated using 5000 iterations of a Markov Chain Monte Carlo (MCMC).The estimate of μ_i with the highest posterior density are selected and the lab-wise normalized $\tilde{\mu}_{il}$ are computed as $\tilde{\mu}_{il} = \frac{\mu_i - u_{\mu l}}{\sigma_{\mu l}}$. The estimated $\tilde{\mu_{il}}$ and $\bar{\tilde{\mu}}_{il}$, the average gregariousness of individual i's peers, are included in the previous model of peer effects as additional variables:

$$Y_{il} = \alpha_l + X_{il}\gamma + \beta\sum_{j\neq i} g_{ij}Y_{jl} + \theta\sum_{j\neq i} g_{ij}X_{jl} + \tilde{\mu}_{il}\tau + \bar{\tilde{\mu}}_{il}\bar{\tau} + \epsilon_{il},\ i,j = 1,\ldots,n,\ l = 1,\ldots,10. \tag{6}$$

Then, $\tau \neq 0$ or $\bar{\tau} \neq 0$ indicates the presence of network endogeneity. Note that the standard error are slightly underestimated as this specification assumes that μ_{il} are observed instead of simulated via MCMC. Houndetoungan (2020) suggests to correct the variance via a Monte Carlo procedure, and this model is thus estimated via 1000 simulations. Moreover, we infer the results in the next section assuming that using labs' fixed effects and accounting for network endogeneity via the inclusion of the estimated gregariousness lead to exogenous errors terms, i.e. the correlated effects' issue is solved and $E(\epsilon_{il}|\alpha_l, X_l, G)=0$. However, the errors might still be heteroskedastic and correlated between workers of the same lab, thus the standard errors of the following estimations are clustered-robust.
This final linear-in-means model of peer effects allows us to estimate causal endogenous peer effects, controlling for exogenous and correlated effects as well as the endogenous network formation. Doing so, we provide evidence concerning which dimensions of social relationships have an influence on individual behavior.

5 Results

5.1 Linear-in-means models of peer effects

When estimating the most basic version of the linear-in-means model (equation 1), we find a significant and strong endogenous peer effect of 0.56 on active transportation mode (Table 4). These effects are of similar magnitude as the endogenous peer effects found by Boucher and Bramoullé (2020) on smoking behavior or by Fletcher (2012) on students' drinking behavior, both using a similar linear-in-means model of peer effects for binary outcomes. One of the advantage of the linear-in-means model for binary outcomes is that the estimated coefficients of the socio-demographics characteristics or endogenous peers effect can be directly interpreted as marginal effects on the likelihood to use an active transportation mode. Thus, having only peers who commute biking or walking increases the likelihood to commute with a similar mode by about 50 percentage points compared to an individual who does not have any peers using an active transportation mode. More realistically, an increase of 20 percentage points in the fraction of peers (one out of 5 peers) who use an active transportation modes results in a 10 percentage points increase in the own probability of choosing such transportation mode. The endogenous peer effects are smaller than 1 in the different models, which is the condition to have a unique solution in the linear-in-means model of Boucher and Bramoullé (2020). The estimated peer effects are also close to the effects found in a study of carpooling and commuting alone in the US military, which uncover that an increase of 20 percentage points in the fraction of peers that carpool increases the probability of carpooling by 10 percentage points (Morrison and Lin Lawell, 2016).

Table 4: Linear-in-means models of peer effects on Active Transportation's choice

	En. effect only	En. + Ex. effects	En., Ex.+ FE effects	All effects
Intercept	0.082	−0.233	–	–
	(0.206)	(0.237)		
Age	−0.007	−0.005	−0.004	−0.005
	(0.005)	(0.005)	(0.005)	(0.005)
Male	0.056	0.093	0.097	0.094
	(0.054)	(0.063)	(0.064)	(0.060)
Distance	0.000	0.000	0.000	0.001
	(0.003)	(0.002)	(0.002)	(0.002)
Yrs in lab	0.004	0.003	0.003	0.004
	(0.005)	(0.006)	(0.005)	(0.005)
Opportunity Cost of active mode (%)	−0.001**	−0.001***	−0.002***	−0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Number of children under 12yrs	0.009	0.019	0.022	0.025
	(0.018)	(0.025)	(0.024)	(0.027)
PhD	0.046	0.056	0.037	0.024
	(0.031)	(0.041)	(0.056)	(0.059)
Revenue 4000−6000€	0.057	0.012	0.007	0.017
	(0.056)	(0.061)	(0.056)	(0.057)
Revenue less 2000€	0.044	0.042	0.038	0.038
	(0.089)	(0.095)	(0.098)	(0.096)
Revenue more 6000€	0.214*	0.185*	0.209	0.211
	(0.103)	(0.092)	(0.118)	(0.119)
Health Issue	−0.188*	−0.211*	−0.256*	−0.266*
	(0.085)	(0.090)	(0.111)	(0.116)
Environmental Sensibility Index	0.372*	0.484**	0.513**	0.515**
	(0.167)	(0.194)	(0.187)	(0.181)
Endogenous Peer Effect	0.558**	0.506	0.748**	0.717**
	(0.171)	(0.343)	(0.251)	(0.262)
G_Age	–	0.008	0.015*	0.015**
		(0.005)	(0.007)	(0.006)
G_Male	–	0.020	−0.011	−0.011
		(0.095)	(0.151)	(0.153)
G_Distance	–	−0.007	−0.003	−0.003
		(0.005)	(0.006)	(0.006)
G_Yrs in lab	–	−0.008	−0.012	−0.012
		(0.007)	(0.007)	(0.006)
G_Opportunity Cost of active mode (%)	–	0.001	0.001	0.001
		(0.002)	(0.001)	(0.001)
G_Number of children under 12yrs	–	0.031	0.038	0.033
		(0.071)	(0.071)	(0.069)
G_PhD	–	−0.031	−0.090	−0.065
		(0.112)	(0.146)	(0.144)
G_Revenue 4000−6000€	–	0.003	−0.048	−0.048
		(0.080)	(0.105)	(0.098)
G_Revenue less 2000€	–	0.129	0.105	0.103
		(0.171)	(0.162)	(0.163)
G_Revenue more 6000€	–	−0.022	−0.009	−0.010
		(0.161)	(0.129)	(0.130)
G_Health Issue	–	0.558**	0.583**	0.600**
		(0.161)	(0.202)	(0.186)
G_Environmental Sensibility Index	–	−0.081	−0.127	−0.098
		(0.281)	(0.292)	(0.272)
Labs' Fixed Effects	–	–	X	X
Gregariousness	–	–	–	0.042
				(0.037)
G_Gregariousness	–	–	–	−0.016
				(0.065)
Nb. Observations	334	334	334	334
R^2	0.18	0.23	0.14	0.16
Weak IV test p.value	0.000	0.000	0.000	0.000
Wu-Hausman test p.value	0.150	0.498	0.05	0.061
Overindentification test p.value	0.443	0.320	0.613	0.552
Fitted values inside [0,1]	93%	92%	90%	90%

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4 also gathers the results from the more complete specifications of the linear-in-means model. It clearly appears that adding contextual effects (exogenous peer effects, "En. + Ex. effects", second column, equation 2) results in non-significant endogenous peer effects in active transportation choice. However, most of the peers' characteristics are not significant, with the exception of age and the variable related to health issues. This may indicate that endogenous and exogenous peer effects are positively correlated. Thus, analyses of endogenous peer effects that fail to control for exogenous peer effects present upward biased estimates of peer effects.
On the contrary, accounting for labs' fixed effects (common environment, "En, Ex. + FE effects", third column, equation 3) increases the size and significance of the endogenous peer effects on active transportation mode's choice. Failing to account for common environment through labs' fixed effect thus yields downward biased endogenous peer effects.
Accounting for the network formation process yields similar results, even if the coefficient of the endogenous peer effect is slightly lower, indicating that gregariousness is positively correlated with the peer effects and that estimations of the endogenous peer effects are upward biased if gregariousness is not included. However, as the coefficient for the estimated gregariousness is not significant, network endogeneity does not appear as an issue in our estimation.

Most of the socio-demographic variables are not significant, with the exception of the dummy variable indicating health issue and the environmental sensibility index. Indeed, individuals that have an health issue limiting physical activities are less likely to use an active transportation mode and the individuals that have a strong environmental sensibility are more likely to use an active transportation mode. Moreover, we observe a negative and significant influence of the relative opportunity cost of the active mode compared to car commuting: the higher the opportunity cost (in time) of using an active transportation mode, the less likely an individual will commute biking or walking.

The null hypothesis for the weak instruments tests, i.e. that the instruments are weak, is rejected at a 1% significance level in all models of peer effects which indicates that the instruments are jointly strong. The null hypothesis of the Wu-Hausman tests, i.e. a two stage least squares (2SLS) modeling is as consistent as an ordinary least squares (OLS) is rejected at a 10% level for our preferred model (the most complete model, "All effects", Table 4), indicating that the 2SLS model is more consistent due to the presence of endogeneity in peers' transportation mode. The null hypothesis of the overindentification tests (Wooldrigde's robust score test, an extension of the Sargan test for clustered-robust standard errors) is not rejected here, thus the instruments are jointly valid, i.e. not correlated with the errors.

Moreover, following the recommendation of Boucher and Bramoullé (2020), we verify that the linear-in-means specification is an adequate structural model for a binary outcome such as active transportation choice. Indeed, in the linear-in-means specification, the estimated outcomes are interpreted as probabilities (as in a linear probability model) and must thus lie in $[0, 1]$. In our models, 90 to 93% of the estimated outcomes are indeed predicted in the right bounds, thus confirming that the linear-in-means modelling of social influence on transportation choice is adequate. Especially, the most basic model of peers effects ("En. effect only") has 93% of its predictions in $[0, 1]$ and the complete model of peer effects ("All effects") reaches 90%.

5.2 Alternative specifications and robustness

5.2.1 Socio-demographic variables

To assess the robustness of our results, we also test the inclusions of several socio-demographic variables or specifications (quadratic, interactions) in the first model (equation 1). The results hold with different specifications of the socio-demographic variables (Appendix, Table A.1). The only notable alternative specification is the one removing the environmental sensibility index from the model. Doing so increases the coefficient for the endogenous peer effects by 17%, indicating that the strength of social influence and pro-environmental values are positively correlated.

5.2.2 Sustainable transportation mode as outcome

In order to check if the empirical strategy is robust to a different specification of the dependent variable, active transportation mode's choice, we estimate the models using the choice of a sustainable transportation mode. Sustainable transportation modes aggregate active modes and public transportation and are chosen by 76% of our sample. The most basic model of peer effects is very similar when analysing sustainable transportation mode's choice (Table A.7, "En. effect only"). The second model, including exogenous peer effects also yields a non significant endogenous peer effect, confirming that endogenous and exogenous peer effects are positively correlated. However, when labs' fixed effects and gregariousness are accounted for, the endogenous peer effects are smaller and non significant, which contrasts with the results of active transportation mode's choice. Yet, the dilution of peer effects might be caused by the high share of respondents adopting sustainable transportation in our sample.

5.2.3 Exclusion of the largest labs

The results are also robust to the exclusion of the largest or the three largest labs, which yields qualitatively similar coefficients although the standard errors increase due to the lower sample size (Table A.2).

5.2.4 Weighted social interaction matrices

Similarly to Lin (2014), we also estimate the models with different specifications of the social interaction matrix representing the networks. Firstly, we test a weighted G, where g_{ij} is not binary (0 or 1) but continuous, representing the strength of the relationship. The strength of the social relationship is constructed from specific questions gathered during the survey: do i and j see themselves outside work, do they eat together, take coffee breaks together or work together. Weighting the social interaction matrix weakens the significance and value of the endogenous peer effects' coefficients in all models (Table A.3).

5.2.5 Heterogeneous peer effects

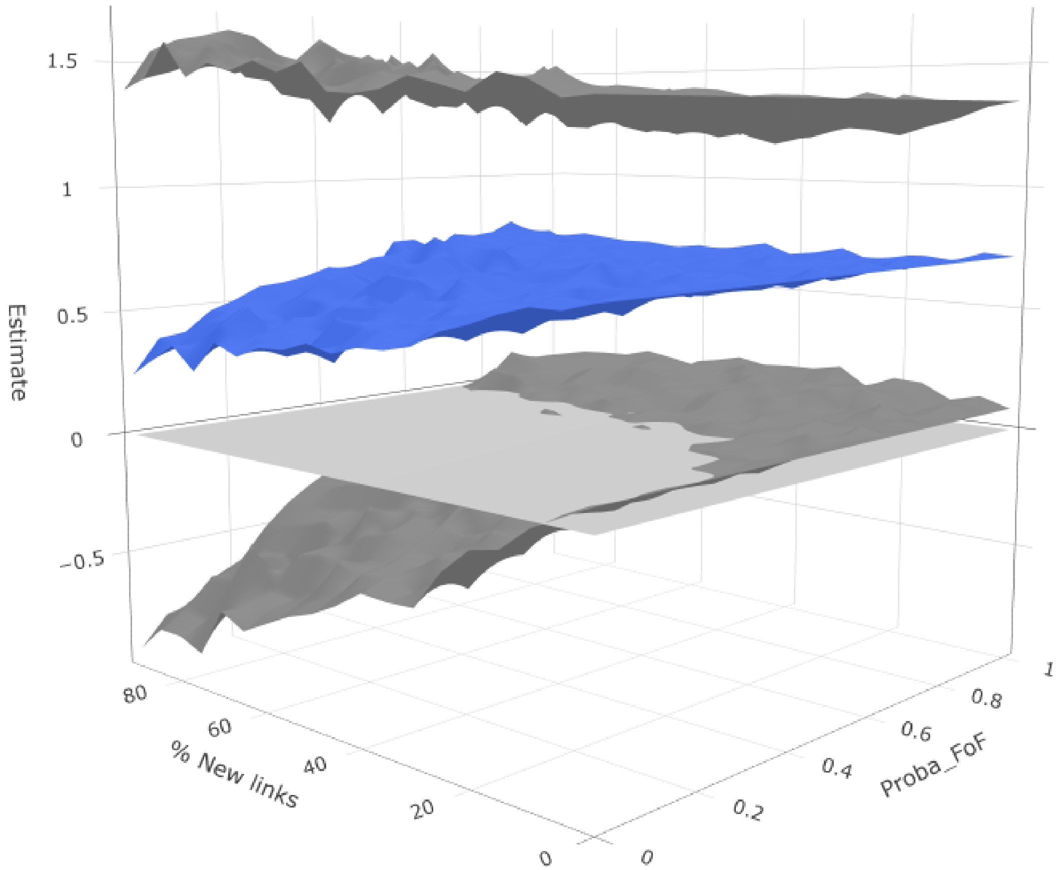
An alternative specification of the heterogeneity of social relationships and peer effects, i.e. the fact that some peers have more influence than others in one's transportation mode, is also analysed building on Dieye and Fortin (2017) and Beugnot et al. (2019). We hypothesize that peer effects in transportation mode are heterogeneous: an individual living close to her workplace will not be socially influenced by her peers living far from work and vice-versa. We decompose the social interaction matrix G in four matrices, G_{YY}, G_{YN}, G_{NY} and G_{NN} where the Y subscript means that the individual lives nearby work (at less than 5km) and N indicates that he/she lives far. Thus, in G_{YY}, $g_{ij} = 1$ only if i and j are peers and both are living close to work and 0 otherwise. In light of previous results (Phithakkitnukoon et al., 2017), we expect that the peer effects between two individuals living either both close (G_{YY}) or both far (G_{NN}) from work will be strong, as they share similar commuting constraints, whereas the peer effects associated with G_{YN} and G_{NY} will be negligible. We also test the hypothesis that individuals are heterogeneously affected by peer effects depending on them having young children or not. Indeed, an individual having young children (less than 12 years old) may be more influenced by her peers' transportation choices if they share similar constraints such as bringing or picking up their children from school.
However, our analyses show that accounting for heterogeneous peer effects based on commuting distance or having young children yields non-significant peer effects (Table A.4 and A.5). This tends to show that peer effects at the workplace are not heterogeneous in the case of transportation mode choice, although this specification likely suffers from false negative bias due to the small samples resulting from the decomposition of the social interaction matrices.

5.2.6 Missing links

In addition, we adapt a method introduced by Patacchini et al. (2017) to account for possible missing links in the adjacency matrix. Indeed, as explained in Data collection and descriptive statistics, participants in the study could nominate six peers maximum. However, 15% of participants cited six peers and we could expect that at least some of them would have nominated more peers if they could. In a similar way, some participants might be reluctant to give some of their colleagues' names to unknown researchers. Thus, some links are likely to be missing in our estimation and may bias the results. To assess the existence and direction of these biases, we simulate adjacency matrices with a varying number of additional links between the individuals.

We make the percentage of new links *(% New links)* vary from 0% to 90% of the observed ("real") links. As these additional links cannot be created between two individuals that are already peers, they are created between individuals who were not directly connected. However, we can force the additional links to be created between an individual and some peer ofher peers or between an individual and another which are not connected by a common peer. The probability that an additional link is created between an individual and a friend of friend *(Proba_FoF)* thus varies from 0 to 1. We make these two variables, *% New links* and *Proba_FoF* increase progressively by a 0.05 percentage points step, thus simulating 360 different adjacency matrix, with a varying amount of missing links. Using each social interaction matrix, we estimate our complete model of peer effects (equation 6). Lastly, we use a bootstrapping strategy and repeat the above process 1000 times, in order to obtain confidence intervals for the endogenous peer effects estimated with the simulated social interaction matrices. Note that for these robustness simulations, we use the estimate of the complete model of peer effects without the Monte Carlo simulations correcting for the two step estimation to reduce the computational burdens.

Figure 1: Peer effects on Active Transportation with missing links



Note: "% New links" (0%-90%) is the percentage of links created while "Proba_FoF" is the probability that additional links are created with a friend of friend rather than with an unrelated worker. The blue surface represents the estimates of the endogenous peer effects while the grey surfaces are the higher and lower bound of the 95% confidence interval. An interactive plot is available at https://chart-studio.plotly.com/~MathieuL/1.embed.

Simulating new social relationships between individuals that are peers' of peers yields virtually identical estimate of endogenous peer effects. Especially, if all new links are created among friends of friends, the endogenous peer effects stay significant for any percentage of simulated links. However, simulating links between unconnected individuals (without common friends) decreases the estimates of peer effects. When 30% or more of the links are created between unconnected individual, the peer effects' coefficients become non-significant for any share of new links. Thus, the fact that these links are formed with peers of peers or unrelated individuals strongly affects the value or significance of the endogenous peer effects.

5.2.7 Bounded number of peers

Similarly, we test the robustness of our results to another data related issue, the fact that most of the respondents could not cite all their peers as some did not agree to participate to the study. As we designed a question in the survey to know if a respondent was able to cite all her peers, we estimate the models with a sub-sample of the 72 participants that could cite all their peers. The coefficients for the endogenous peer effects are similar to our main models in this specification (A.6) although not significant, which is expected due to the small sample size.

5.2.8 Falsification test of the adjacency matrix

Furthermore, we propose a falsification test (An, 2015; De Giorgi et al., 2010). Let imagine that the networks we collect are highly mismeasured and that the peer effects we observed are actually generated by sample variation or measurements errors due to partial networks. Under this assumption, if we construct other adjacency matrices artificially and randomly but with similar density as the observed adjacency matrix, we should also be able to observe similar significant peer effects due to errors-in-variable bias. Conversely, if the peer effects under the artificial and random adjacency matrices are virtually zero and non-significant, we may conclude that the peer effects we observe are relatively robust and may be causally interpreted. Thus, we randomly simulate placebo adjacency matrices and estimate the complete model of peer effects using the related social interaction matrices. In a bootstrap framework, we simulate 1000 different social interaction matrices and thus estimate 1000 times the coefficients to create enough variation.

Satisfactorily, the endogenous and exogenous peer effects are virtually zero and non-significant[2], indicating that the peer effects we observe in our sample are not due to random variation or measurement errors.

5.2.9 Simulated maximum likelihood estimator

Importantly, we implement the estimator of peer effects for binary outcome of Lee et al. (2014). This estimator accounts for heterogeneous rational expectation of peers' outcomes and is thus estimated using a simulated maximum likelihood estimation of a logit specification. In this framework, individuals form different expectations of peers' transportation modes based on these peers' characteristics and select their own transportation mode according to these expectations. In this sense, the relationship between the expectations of peers' behavior and their characteristics is non-linear and allows for identification of endogenous and exogenous peer effects, conversely to the linear-in-means model without instrumentation (Brock and Durlauf, 2001). However, as this estimator is based on a logit specification, the coefficients cannot be directly interpreted as marginal effects as in the linear-in-means model. We thus compute the "naive" marginal effects following Lee et al. (2014). The coefficient of the endogenous peer effects is positive and significant using this estimator although half the size of the coefficient estimated with the linear-in-means model (A.8). The marginal effects associated to the endogenous peer effects are three times smaller when estimated with the model of Lee et al. (2014). Boucher and Bramoullé (2020) also report marginal effects in the linear-in-means model three

[2]the results are available upon request to the authors

times larger than the marginal effects from the Lee et al. (2014)'s paper. This strong difference in the strength of the coefficients and marginal effects may be explained by the fact that 2SLS estimator using peers of peers' characteristics as instruments corrects for the negative exclusion bias (Caeyers and Fafchamps, 2016).

5.2.10 Non-linearity in the linear-in-means model

Lastly, we perform an additional estimation of our preferred linear-in-means estimator, allowing for non-linearity in the endogenous peer effects. We apply the control function of two-stage least squares proposed by Wooldridge (2015). This approach differs to the classical 2SLS estimation, detailed here for model 1 as the second stage is not estimated using the instrumented $\widehat{GY}$, but using the errors from the first stage, ν and the endogenous independent variable GY. Recall that the 2SLS procedure for begin model 1 is:

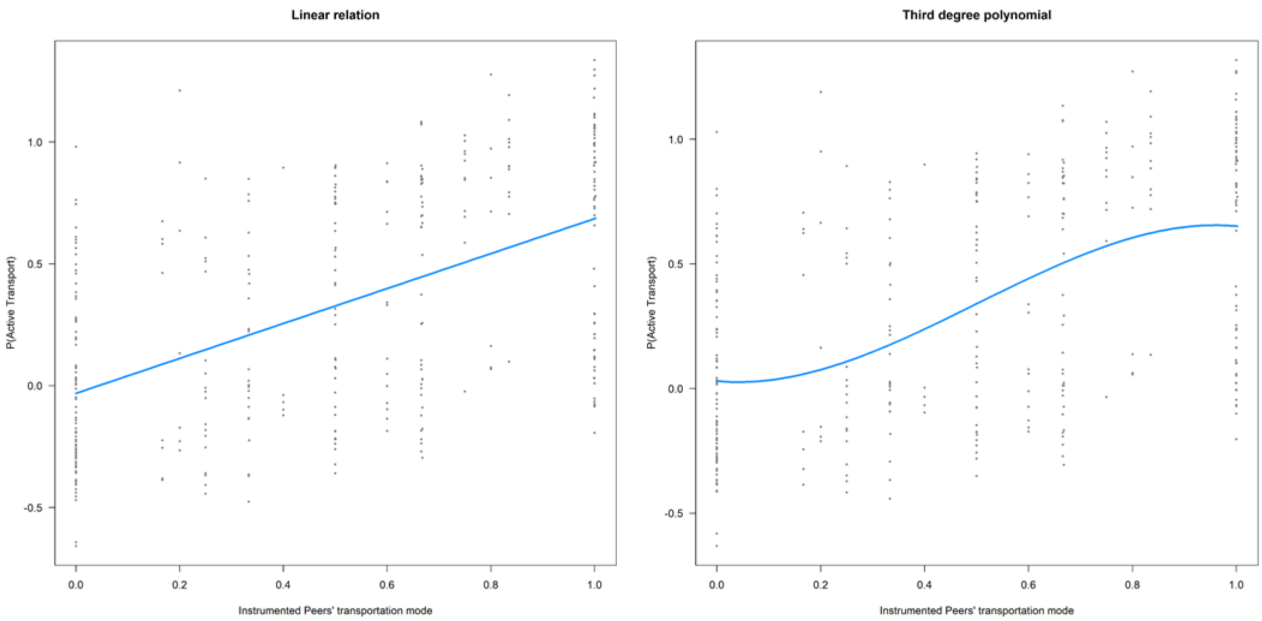
$$Y = \alpha + X\gamma + \beta\widehat{GY} + \epsilon \text{ with } GY = Z\pi + \nu,\ Z = [1, X, G^2X]$$

The control function approach yields:

$$Y = \alpha + X\gamma + \beta GY + \rho\nu + \epsilon \text{ with } GY = Z\pi + \nu,\ Z = [1, X, G^2X]$$

The standard errors are underestimated in this estimator as it does not account for the two-step procedure and we need the additional assumption that both the errors of the first and second stages, ν and ϵ are independent Z. However, the control function approach allows us to estimate non-linear endogenous peer effects by adding to the above-equation with polynomials of degree 2 and 3 of GY (Norris, 2020).

Figure 2: Assessment of the linearity of endogenous peer effects in the complete model



The linear specification appears as a satisfying simplification of the non-linear endogenous peer effects (Figure 2), except at the extreme values of the peers' average active transportation. Indeed, endogenous peer effects are constant in low (0 to 0.2) and high (0.9 to 1) average share of peers who commute with an active transportation mode. Thus, if a given individual has a peer that shifts to an active transportation mode but that the new average peers' transportation behavior is either close to zero or one, then this additional peer commuting with an active mode will not influence the individual. These observations uncover interesting aspects of behavior and social norms: if the behavior in question is too much of a minority in peers' behavior it will not be the social norm an individual would wish to conform to. At the opposite, if the behavior is too much of a majority, conforming to the norm will not yield additional utility.

6 Using peer effects to increase policy effects : selecting key players

Recent meta-analyses have shown that *soft* transport policy measures are effective in reducing car use (Möser and Bamberg, 2008; Semenescu et al., 2020). Such soft interventions are mainly personal travel planning, which include personal coaching sessions to increase self-efficacy, capability, knowledge, awareness about alternative transportation modes or to discuss the role of habits and social, cultural or moral norms (Mathy et al., 2020). However, this type of personalized interventions are difficult to scale (coaches specialized in transportation mode's choice are in limited supply) and costly (90$ to 200$ per employee on average in the UK, USA and Netherlands (Cairns et al., 2008)). Thus, personalized interventions may only be realistically proposed to a small share of the workers, which we here fix at 10%.

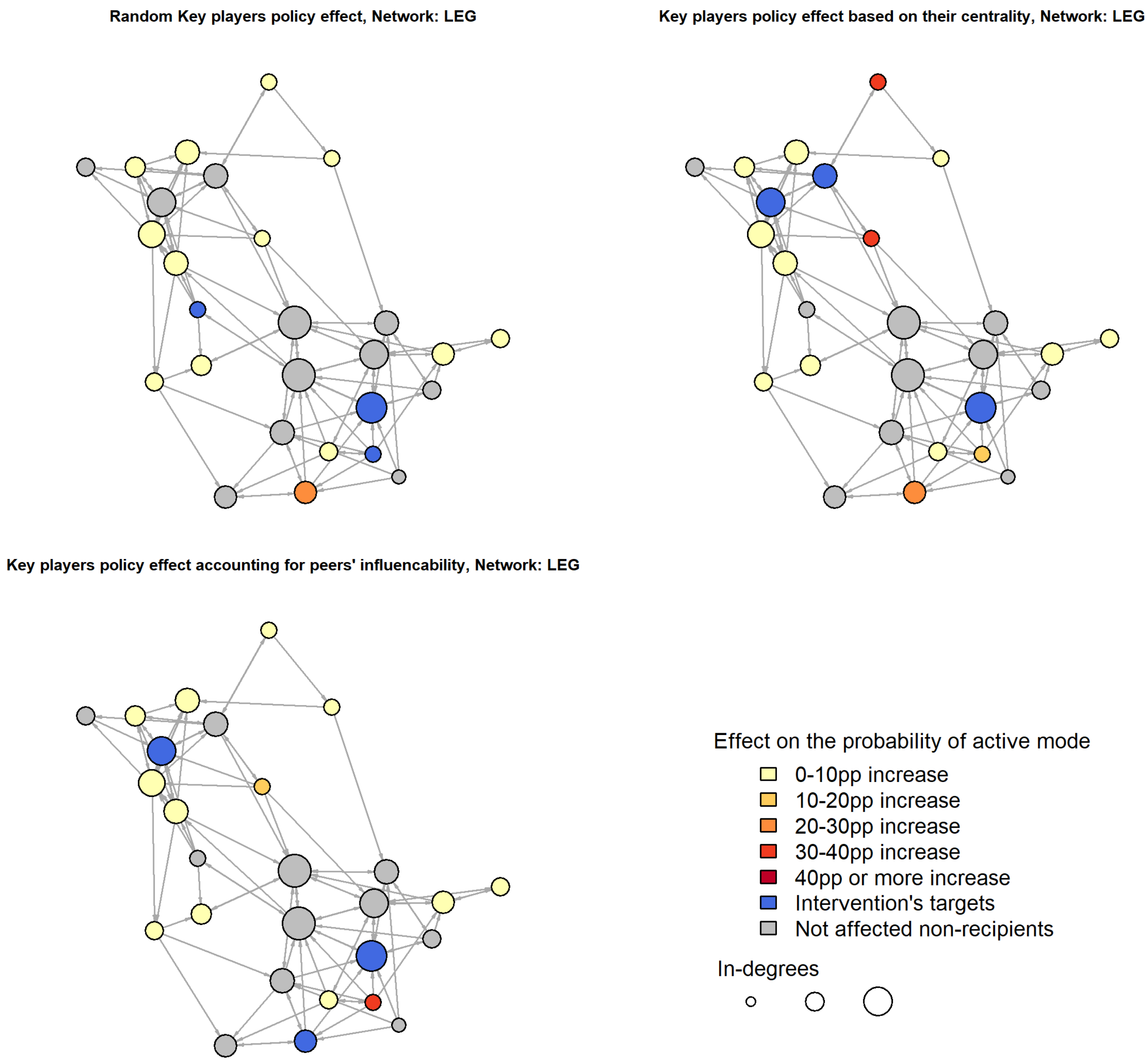
In order to measure the spillover effect emerging from social influence, we simulate the impact of a public policy which would select 10% of the interviewees in each lab (which amounts to 34 individuals, ranging from 1 to 9 recipients per lab), on the realistic conditions that they do not already use an active transportation mode and live 15km or less away from their workplace, which amounts to 116 individuals. The intervention we have in mind relates to personal coaching of employees to raise awareness of the advantages of active mobility and define the best way to commute using a bicycle. As these kind of interventions are costly, an employer or public authority may be willing to pay for only a few coaching sessions. We assume that the policy is perfectly effective, and that the 34 interviewees effectively changed their transportation mode to an active mode. We then compute the spillover impact of this policy, estimating to which extent the selected individuals will affect the transportation mode's choice of the non-selected individuals, through peer effects. The spillover effects is also restricted to the sub-sample of individuals that do not already commute with an active transportation mode and who live 15km or less away from their workplace (Phithakkitnukoon et al., 2017; Pike and Lubell, 2018), i.e. 82 individuals. Thus, the spillover effects we observe are realistic as non-beneficiaries that already use an active transportation mode or live too far from work to credibly shift to an active mode are not accounted for. We consider three different recipients' selection strategies:

- **Random selection.** The beneficiaries of the intervention are randomly selected among eligible individuals. To obtain an averaged estimate of the spillover effect, we repeat 100 times the analysis in this case.

- **Key player selection - based on in-degree centrality.** The key players are the workers with the higher in-degree centrality in each network (lab), i.e. the higher number of individuals who consider them as peers. The in-degree of an individual (or node) i is $deg^{-}(i) = \sum_j a_{ji}$. The selected individuals can thus be considered as key players, as they are selected based on a objective criteria of centrality. This key players' selection aims at maximizing the number of individuals influenced by the key players' changes to an active transportation mode.

- **Key player selection - based on peers' influenceability.** We can further refine the selection of key players by considering not only the number of individuals that cited them as peers, but also the influenceability of these individuals. Indeed, individuals that have one or several key players (as defined by the in-degree centrality criteria) as peer may have a varying amount (from 1 to 6) of peers. However, the peer effects in the linear-in-means model are based on the social group's norm, i.e. the average transportation mode of one's peers. Thus, an individual which has only two peers, among which a key player, selected by the public policy, will be more impacted by the spillover effect of the policy that an individuals with four peers. To account for influenceability, we also compute the in-degree of each node, but in contrast to the previous selection strategy, we use the social interaction matrix. Recalling that the in-degree centrality of a node i is the sum of the i^{th} column of the adjacency matrix A, $deg^{-}(i) = \sum_j a_{ji}$ and that row-normalization implies that if j has five peers, the (non-zero) values in the j^{th} social interaction matrix will be 0.2 instead of 1. Thus, we use here a modified in-degree centrality measure, $deg^{*-}(i) = \sum_j g_{ji}$. If i is recognized as a peer by n individuals which have relatively few peers, then the sum of the i^{th} column of the adjacency matrix will be larger that the one of an alternative individual who is recognized as a peer by n individuals with relatively a lot of peers. Thus, computing in-degree centrality using the social interaction matrix accounts for peers' influenceability and increases the effective change of the social norms in each network compared to classic in-degree selection of key players.

Figure 3: Comparison of spillovers with different selection strategies, LEG lab



Note: The graphs for the nine other labs are available in the Appendix. Not affected non-recipients of the intervention are either already using an active transportation mode or living 15km or more away from their workplace.

Our analysis of spillover effects shows that public policies or interventions that target key players as recipients would be twice as effective as policies that randomly select the recipients. Indeed, when policy's recipients are randomly selected, the share of non-recipients that use an active transportation mode increases from 43.7% to 46.3% (using the estimated parameters from the complete model of peer effects, Table 4). Thus, the spillover effects of the intervention targeting 10% of the workers amount to 5.8% among non-recipients. However, when key players are targeted, uses of the active transportation reach 47.7% and 47.6% among the non-recipients, for key players' selection based on in-degree centrality and influenceability respectively. The spillover effects thus reach 8.1% and 7.8% respectively. Figure 3 illustrates the random (top left quadrant) and the two different key players' selection strategies (top right and bottom left quadrants) and their effect on the spreading of active transportation mode on a given lab. The alternative way of selecting key players, accounting for the influenceability of their peers, shows a slightly lower spillover effect, as one of the targeted key players differs between the two specifications. The graphs for the other nine labs are available in the Appendix (Figures A.2 to A.9).

As data collection for network analysis is generally costly, identifying the key-players through network characteristics (degrees, influence) might be infeasible in some cases. We thus test if socio-demographic characteristics could be good predictors of the probability to be a key-player, using the information on the key-players we selected previously. However, the results[3] show that only age and household revenue are significant predictors and might be specific to our research labs' context.

[3]available upon request to the authors

7 Discussion and conclusion

Our results yield important information on the role of social interactions in adopting an active transportation mode. Firstly, we show that social relationships at the workplace do have a strong influence on individual behavior outside the workplace. An increase of 20 percentage points in the fraction of peers using an active transportation mode of a given individual results in a 10 to 14 percentage points increase in its own probability of also choosing an active transportation mode.

We also show that network analysis may be key in economic evaluation of public policies and interventions as it allows the identification of key players. Indeed, selected key players as the recipients of a public policy increases by 40% the social spillovers of the policy. A policy framing 10% of the population as recipient, thus making them shift from non-active to active transportation mode, would results in a 5.8 to 8.1% increase in the probability of using an active mode of transportation for non-recipients. These spillover effects would result in shifts of transportation modes from car to bicycle or walk. We quantify the savings in *external costs* (pollution, road wearing and accidents) due to the lower car use and in *medical treatment costs* (reduction in cardio-vascular diseases) due to the higher physical activity. The spillover effects of the hypothetical policy targeting key players would results in 5,476€[4] saved annually in *external costs* and 79,091€[5] saved in *medical treatment costs* in the long term.

The methods used to select key players, based on in-degree centrality or peers' influenceability show similar results. This indicates that selecting recipients that maximize the number of non-recipients influenced or selecting recipients that maximize the change of the social norms (defined here as the average behavior on one's peers) in each peer networks are equivalent. However, this simulation of spillover effects does not integrate second order spillover effects, i.e. the facts that the non-recipients influenced by the selected key players may also influence other non-recipients if they also shift to an active transportation mode. Also, we do not take into account the fact that the recipients (or non-recipients) were previously using either a car or public transportation to commute. However, this difference might influence the effectiveness of the public policy, as it may be smoother and easier to shift to an active transportation mode from public transportation than from private car.

We also provide important evidence that the linear-in-means model of peer effects is an adequate modelling strategy in the presence of binary outcome, active transportation mode's

[4]The *external costs* are estimated as 0.174€/km (Bergerot et al., 2021), and the savings due to the spillover effects are computed using the average commuting trip's distance in the sample and the baseline share of car users (24%), assuming that the individuals who shifts to an active transportation mode will not use a car to commute anymore and that they commute 5 days a week, 43 weeks a year.

[5]The *medical treatment costs* are estimated as 138,000€ per cardio-vascular disease (Bouscasse et al., 2022) and the savings due to the spillover effects are computed assuming that cardio-vascular diseases affect 30% of the population and that shifts to an active transportation reduces by 5 percentage points the risk to be affected by such diseases (Rojas-Rueda et al., 2013).

choice and in the professional context. Indeed, to the best of our knowledge, the linear-in-means model has been apply only twice, on alcohol consumption (Fletcher, 2012) or smoking (Boucher and Bramoullé, 2020), both time using the "Add Health" data and thus friendship social relationships. We find endogenous peer effects of similar size and significance than the above-cited articles, even though we use a smaller and less exhaustive data set and coworkers networks, which tends to show that the linear-in-means model with instrumental variables is an interesting estimator of endogenous peer effects when collection of network data is costly or non-exhaustive. Moreover, we apply the simulated maximum likelihood estimator of peer effects accounting for heterogeneous rational expectations of Lee et al. (2014) and our results are supporting the findings of Boucher and Bramoullé (2020): the estimated coefficients are similar in both models but the marginal effects of Lee et al. (2014)'s estimator, which is based on a logit specification, are three times smaller than the marginal effects of the equivalent linear-in-means model.

In the case of active transportation mode, accounting for exogenous peer effects and correlated effects helps identifying a stronger and more significant endogenous peer effects. The results are robust to different specifications of independent variables or to the weighting of the social interaction matrix by the intensity of the social relationships. However, accounting for the heterogeneity of peers effects based on spatial or familial constraints does not yield significant peer effects. Moreover, modifying the depend variable to include public transportation reduces the size and significance of peer effects. This dilution of the endogenous peer effects may be caused by the fact that a majority (76%) of respondents in our study use a sustainable transportation mode while only 44% use an active transportation mode. Identifying peer effects is intuitively more difficult when the behavior in question is already the norm in the social groups under scrutiny. In addition, social influence and peer effects might be stronger for active transportation mode, as they are related to sportive motivation and (friendly) competition which may create a stronger social emulation than public transportation. However, public transportation might be preferred by individuals with higher gregariousness as this transportation mode facilitate social interactions. Pike (2014) also hints that peers effects are more relevant to active transportation mode while neighborhood (spatial and socio-economic) effects are more relevant to public transportation.

Nevertheless, these results must be cautiously considered as we faced some empirical challenges. Firstly, our data set is rather small, compared with other applications of social network to economic topics, which usually have access to thousands of individuals and larger networks. A small data set increases the risk of type II errors and more socio-demographic variables could be found significant in a larger study. Secondly, the data set we have is not randomized and a selection bias may arise as the individuals that respond to the survey might be the most socially integrated in the research labs and thus might have a denser social network or a higher propensity to conform to social norms. Thirdly, the data we collected on network is partial. We do not observe the full network as, on average, only 35% of the labs' personnel answered the survey. As these 334 participants could only nominate peers who also accepted to participate in the study, there is certainly a substantial number of missing nodes and edges in the induced networks. While our robustness analyses account for possible missing edges and measurement errors in the networks, the biases of the estimated coefficients due to the missing nodes remain (Chandrasekhar and Lewis, 2016). However, the sample we use is not representative of the French population, as only 25% of the respondents to our study use private cars to commute, whereas the average in France is at least two to three times as large (Brutel and Pages, 2021). Nonetheless, the average transportation modes of our sample are in line with the specific situation of Grenoble, which is one of the cities where inhabitants uses the most active and sustainable transportation in France, because of political (environmentalist mayor) and topographic (the city is extraordinary flat) reasons (Syndicat Mixte des Mobilités de l'Aire Grenobloise, 2021). Similarly, we surveyed research labs, among which the workers are generally more educated, have a higher revenue and are more aware of the environmental and health externalities of the transportation modes than the average French worker. Social relationships among research labs could also differ in nature in private companies where hierarchical relationships may play a larger role. The strong peer effects we estimate may thus not generalized to the whole population nor to other cities and further research is necessary to externally validate our results.

Conducting a larger survey using the recently developed data collection strategy of Breza et al. (2020), using *Aggregated Relational Data* could resolve some of these empirical issues. Similarly, gathering panel data or information on a larger variety of trips per individual - not only commuting - might refine our results (Kim et al., 2017). Especially, individuals plan their trips as a tour, and commuting trips are not always independent of trips planned after or before work. Moreover, gathering panel data can reveal additional information, such as the estimation of long-term peer effects after a policy or intervention (Alacevich et al., 2021) or the dynamic evolution of the network when individuals change their behaviors (Badev, 2021; Boucher, 2016). Collecting panel data on the network would also allow to observe and not simulate the marginal impact of link formation or dissolution on individual behavior as well as the impact of individual behavior on these processes.

8 References

Akerlof, G. (1997), 'Social distance and social decisions', *Econometrica* **65**(95), 1005.

Alacevich, C., Bonev, P. and Söderberg, M. (2021), 'Pro-environmental interventions and behavioral spillovers: Evidence from organic waste sorting in sweden', *Journal of Environmental Economics and Management* **108**.

An, W. (2015), 'Instrumental variables estimates of peer effects in social networks', *Social Science Research* **50**, 382–394.

Badev, A. (2021), 'Nash equilibria on (un)stable networks', *Econometrica* **89**(3), 1179–1206.

Bergerot, A., Comolet, G. and Salez, T. (2021), 'Les usagers de la route paient-ils le juste prix de leurs circulations ?', *Trésor-Eco* **283**.

Beugnot, J., Fortin, B., Lacroix, G. and Villeval, M. C. (2019), 'Gender and peer effects on performance in social networks', *European Economic Review* **113**, 207–224.

Boucher, V. (2016), 'Conformism and self-selection in social networks', *Journal of Public Economics* **136**, 30–44.

Boucher, V. and Bramoullé, Y. (2020), 'Binary outcomes and linear interactions', *AMSE Working Papers 2038* .

Boucher, V., Bramoullé, Y., Djebbari, H. and Fortin, B. (2014), 'Do peers affect student achievement? evidence from canada using group size variation', *Journal of Applied Econometrics* **29**, 91–109.

Bouscasse, H., Gabet, S., Kerneis, G., Provent, A., Rieux, C., Ben Salem, N., Dupont, H., Troude, F., Mathy, S. and Slama, R. (2022), 'Designing local air pollution policies focusing on mobility and heating to avoid a targeted number of pollution-related deaths: Forward and backward approaches combining air pollution modeling, health impact assessment and cost-benefit analysis', *Environment International* **159**.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* **150**, 41–55.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2019), 'Peer effects in networks: a survey', *Annual Review of Economics* **12**, 603–619.

Breza, E., Chandrasekhar, A. G., McCormick, T. H. and Pan, M. (2020), 'Using Aggregated Relational Data to Feasibly Identify Network Structure without Network Data', *American Economic Review* **110**(8), 2454–2484.

Brock, W. A. and Durlauf, S. N. (2001), 'Discrete Choice with Social Interactions', *The Review of Economic Studies* **68**(2), 235–260.

Brutel, C. and Pages, J. (2021), La voiture reste majoritaire pour les déplacements domicile-travail, même pour de courtes distances, Technical Report 1835.

Caeyers, B. and Fafchamps, M. (2016), ‘Exclusion bias in the estimation of peer effects’, (22565).

Cairns, S., Sloman, L., Newson, C., Anable, J., Kirkbride, A. and Goodwin, P. (2008), ‘Smarter choices: Assessing the potential to achieve traffic reduction using ‘soft measures”, *Transport Reviews* **28**(5), 593–618.

Calvo-Armengol, A., Patacchini, E. and Znou, Y. (2009), ‘Peer effects and social networks in education’, *Review of Economic Studies* **76**(4), 1239–1267.

Chandrasekhar, A. G. and Lewis, R. (2016), ‘Econometrics of sampled networks’.

Cialdini, R. B. and Goldstein, N. J. (2004), ‘Social influence: Compliance and conformity’, *Annu. Rev. Psychol.* **55**, 591–621.

De Giorgi, G., Pellizzari, M. and Redaelli, S. (2010), ‘Identification of social interactions through partially overlapping peer groups’, *American Economic Journal: Applied Economics* **2**(2), 241–75.

De Witte, A., Hollevoet, J., Dobruszkes, F., Hubert, M. and Macharis, C. (2013), ‘Linking modal choice to motility: A comprehensive review’, *Transportation Research Part A: Policy and Practice* **49**, 329–341.

Dieye, R. and Fortin, B. (2017), ‘Gender peer effects heterogeneity in obesity’, *SSRN Journal* .

Dugundji, E. and Gulyás, L. (2008), ‘Sociodynamic discrete choice on networks in space: impacts of agent heterogeneity on emergent outcomes’, *Environ. Plann. B: Plann. Des.* **35**(6), 1028–1054.

Fletcher, J. (2012), ‘Peer influences on adolescent alcohol consumption: evidence using an instrumental variables/fixed effect approach’, *Journal of Population Economics* **25**(4), 1265–1286.

Fortin, B. and Yazbeck, M. (2015), ‘Peer effects, fast food consumption and adolescent weight gain’, *Journal of Health Economics* **42**, 125–138.

Goetzke, F. and Andrade, P. (2010), ‘Walkability as a summary measure in a spatially autoregressive mode choice model: An instrumental variable approach’, *Spatial Statistics Methods and Applications* .

Goetzke, F. and Weinberger, R. (2012), ‘Seperating contextual effects from endogenous effects in automobile ownership models’, *Environment and Planning A* **44**, 1032–1046.

Graham, B. S. (2017), ‘An Econometric Model of Network Formation With Degree Heterogeneity’, *Econometrica* **85**(4), 1033–1063.

Houndetoungan, E. A. (2020), 'A Count Data Model with Social Interactions', *SSRN Electronic Journal* .

Hsieh, C.-S., Lin, X. and Patacchini, E. (2020), Social Interaction Methods, *in* K. F. Zimmermann, ed., 'Handbook of Labor, Human Resources and Population Economics', Springer International Publishing, Cham, pp. 1–30.

IPSOS and BCG (2017), Observatoire européen des mobilités, Technical report.

Kelejian, H. H. and Prucha, I. R. (1998), 'A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances', *The Journal of Real Estate Finance and Economics* **17**, 99–121.

Kim, C., Parent, O. and vom Hofe, R. (2017), 'The role of peer effects and the built environment on individual travel behavior', *Environment and Planning B: Urban Analytics and City Science* **45**, 452–469.

Kormos, C., Gifford, R. and Brown, E. (2015), 'The influence of descriptive social norm information on sustainable transportation behavior: A field experiment', *Environment and Behavior* **47**, 479–501.

Landini, F., Montinari, N. and Piovesan, M. (2016), 'Friendship network in the classroom: Parents bias of peer effects', *Journal of Economic Behavior & Organisation* **129**, 56–73.

Lee, L.-f., Li, J. and Lin, X. (2014), 'Binary Choice Models with Social Network under Heterogeneous Rational Expectations', *Review of Economics and Statistics* **96**(3), 402–417.
URL: *https://direct.mit.edu/rest/article/96/3/402-417/58168*

Lin, X. (2014), 'Peer effects in adolescents' delinquent behaviors: Evidence from a binary choice network model', *Regional Science and Urban Economics* **46**, 73–92.

Lindquist, M. J., Sauermann, J. and Zenou, Y. (2015), 'Network effects on worker productivity', *CEPR Discussion Papers 10928* .

Malcolm Morgan, Marcus Young, Robin Lovelace and Layik Hama (2019), 'Opentripplanner for r', *Journal of Open Source Software* **4**(44), 1926.

Maness, M., Cirillo, C. and Dugundji, E. R. (2015), 'Generalized behavioral framework for choice models of social influence: Behavioral and data concerns in travel behavior', *Journal of Transport Geography* **46**, 137–150.

Manski, F. C. (1993), 'Identification of endogenous social effects: The reflection problem', *Review of Economic Studies* **60**, 531–542.

Mathy, S., Bouscasse, H., Chardonnel, S., Chalabaëv, A., Gabet, S., Treibich, C. and Slama, R. (2020), *Protocol of an Interdisciplinary and Multidimensional Assessment of Pollution Reduction Measures in Urban Areas: MobilAir Project*, Springer International Publishing,

Cham, pp. 325–345.
URL: *https://doi.org/10.1007/978-3-030-30978-7$_1$8*

Morrison, G. M. and Lin Lawell, C.-Y. C. (2016), ‘Driving in force: The influence of workplace peers on commuting decisions on U.S. military bases’, *Journal of Economic Behavior & Organization* **125**(C), 22–40.

Möser, G. and Bamberg, S. (2008), ‘The effectiveness of soft transport policy measures: A critical assessment and meta-analysis of empirical evidence’, *Journal of Environmental Psychology* **28**(1), 10–26.

Norris, J. (2020), ‘Peers, parents, and attitudes about school’, *Journal of Human Capital* **14**(2), 290 – 342.

Patacchini, E., Rainone, E. and Zenou, Y. (2017), ‘Heterogeneous peer effects in education’, *Journal of Economic Behavior & Organization* **134**, 190–227.

Phithakkitnukoon, S., Sukhvibul, T., Demissie, L., Smoreda, Z., Natwichai, J. and Bento, C. (2017), ‘Inferring social influence in transport mode choice using mobile phone data’, *EPJ Data Science* **6**, 11.

Pike, S. (2014), ‘Travel mode choice and social and spatial reference groups: Comparison of two formulations’, *Transportation Research Record* **1**(75-81), 531–542.

Pike, S. and Lubell, M. (2016), ‘Geography and social networks in transportation mode choice’, *Journal of Transport Geography* **57**, 184–193.

Pike, S. and Lubell, M. (2018), ‘The conditional effects of social influence in transportation mode choice’, *Research in Transportation Economics* **68**, 2–10.

Powell, L. M., Tauras, J. A. and Ross, H. (2005), ‘The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior’, *Journal of Health Economics* **24**, 950–968.

Rojas-Rueda, D., de Nazelle, A., Teixidó, O. and Nieuwenhuijsen, M. (2013), ‘Health impact assessment of increasing public transport and cycling use in barcelona: A morbidity and burden of disease approach’, *Preventive Medicine* **57**(5), 573–579.

Semenescu, A., Gavreliuc, A. and Sârbescu, P. (2020), ‘30 years of soft interventions to reduce car use -a systematic review and meta-analysis’, *Transportation Research Part D Transport and Environment* **85**, 1–15.

Sheeran, P. and Webb, T. L. (2016), ‘The Intention-Behavior Gap’, *Social and Personality Psychology Compass* **10**(9), 503–518.

Sherwin, H., Chatterjee, K. and Jain, J. (2014), ‘An exploration of the importance of social influence in the decision to start bicycling in england’, *Transportation Research Part A: Policy and Practice* **68**, 32–45.

Syndicat Mixte des Mobilités de l'Aire Grenobloise (2021), 'Enquête mobilité grande région grenobloise 2020 - premiers résultats'.

Videras, J., Conover, E. and Wu, S. (2011), 'The influence of social relationships on pro-environment behaviors', *Journal of Environmental Economics and Management* **63**.

Walker, J. L., Ehlers, E., Banerjee, I. and Dugundji, E. (2005), 'Correcting for endogeneity in behavioral choice models with social influence variables', *Transportation Research Part A: Policy and Practice* **45**(4), 362–374.

Wang, H., Akar, G. and Guldmann, J. (2015), 'Do your neighbors affect your bicycling choice? a spatial probit model for bicycling to the ohio state university', *J. Transp. Geogr.* **42**, 122–130.

Wooldridge, J. M. (2015), 'Control function methods in applied econometrics', *The Journal of Human Resources* .

9 Appendix

Figure A.1: Graphical representation of the networks

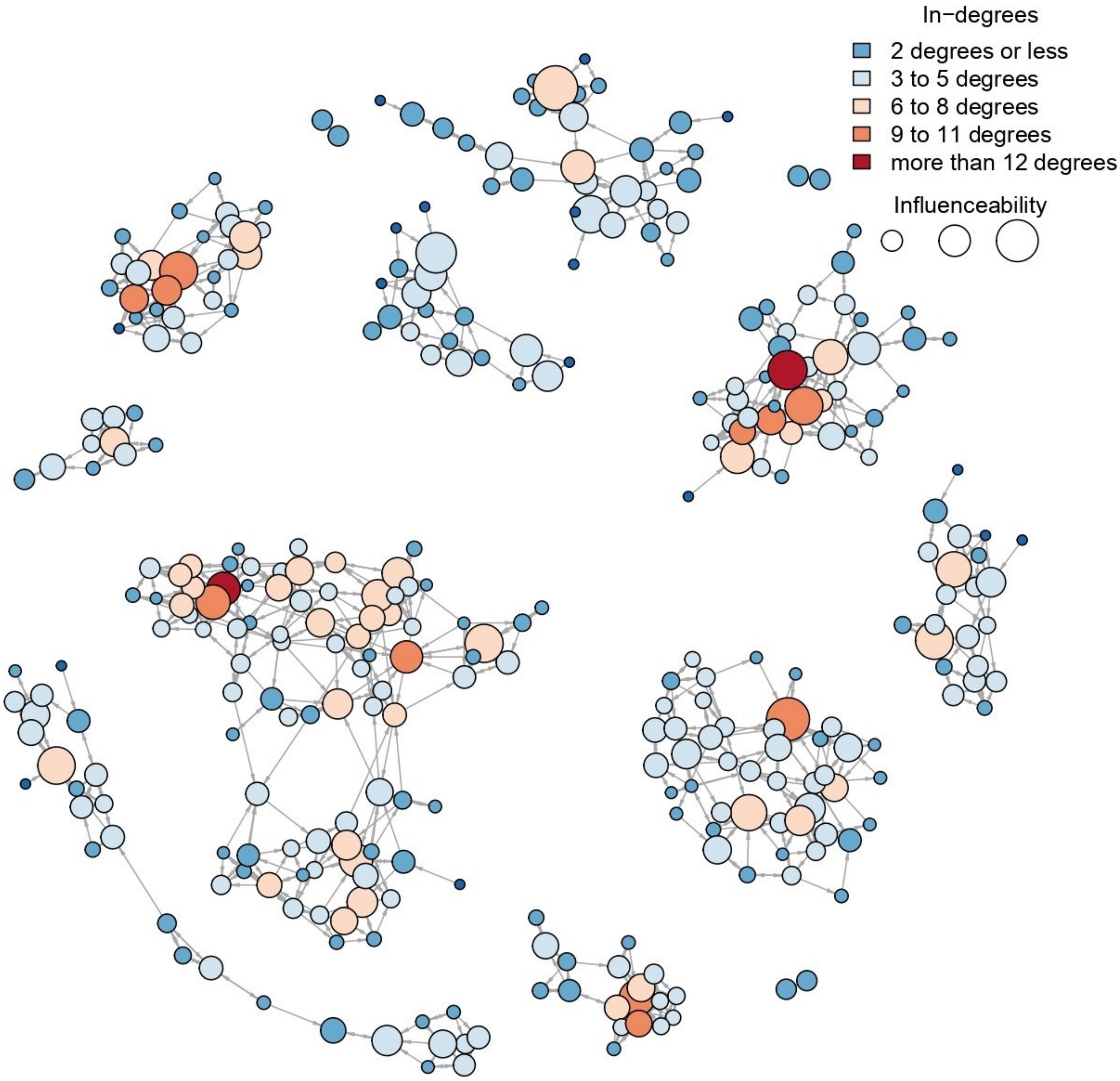


Table A.1: Robustness of different specifications of the model of endogenous peer effects only

	Baseline model	Square age	Square distance	Interaction age:children	All children	Ø health issue	Ø environmental sensibility
Intercept	0.082	−0.384*	0.100	0.087	0.104	0.073	0.331**
	(0.206)	(0.184)	(0.181)	(0.206)	(0.205)	(0.206)	(0.119)
Age	−0.007	0.016	−0.007	−0.008	−0.007	−0.008	−0.008
	(0.005)	(0.013)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)
Male	0.056	0.063	0.066	0.053	0.067	0.061	0.028
	(0.054)	(0.054)	(0.050)	(0.055)	(0.055)	(0.053)	(0.061)
Yrs in lab	0.004	0.006	0.005	0.005	0.003	0.005	0.005
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Distance	0.000	0.000	−0.014*	0.000	0.000	0.000	0.000
	(0.003)	(0.003)	(0.006)	(0.003)	(0.003)	(0.003)	(0.002)
Opportunity Cost of active mode (%)	−0.001**	−0.002**	0.000	−0.001**	−0.002***	−0.001**	−0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of children under 12yrs	0.009	−0.007	0.016	−0.032	–	0.008	0.018
	(0.018)	(0.022)	(0.018)	(0.099)		(0.018)	(0.018)
PhD	0.046	0.025	0.053	0.046	0.051	0.046	0.049
	(0.031)	(0.030)	(0.036)	(0.032)	(0.031)	(0.032)	(0.034)
Revenue 4000−6000€	0.057	0.059	0.035	0.058	0.058	0.055	0.067
	(0.056)	(0.053)	(0.055)	(0.056)	(0.054)	(0.054)	(0.057)
Revenue less 2000€	0.044	0.064	0.028	0.040	0.060	0.040	0.021
	(0.089)	(0.089)	(0.081)	(0.089)	(0.081)	(0.090)	(0.085)
Revenue more 6000€	0.214*	0.215*	0.165	0.213*	0.211*	0.208	0.232*
	(0.103)	(0.102)	(0.098)	(0.104)	(0.093)	(0.111)	(0.104)
Health Issue	−0.188*	−0.188*	−0.176	−0.191*	−0.164*	–	−0.217**
	(0.085)	(0.083)	(0.100)	(0.086)	(0.080)		(0.086)
Environmental Sensibility	0.372*	0.384*	0.459**	0.369*	0.398**	0.375*	–
	(0.167)	(0.163)	(0.157)	(0.167)	(0.158)	(0.172)	
Age^2	–	0.000	–	–	–	–	–
		(0.000)					
$Distance^2$	–	–	0.000	–	–	–	–
			(0.000)				
Interaction age:children under 12yrs	–	–	–	0.001	–	–	–
				(0.002)			
Number of children	–	–	–	–	0.028	–	–
					(0.018)		
Endogenous Peer Effect	0.558**	0.536**	0.597***	0.579**	0.424*	0.567**	0.672**
	(0.171)	(0.173)	(0.155)	(0.177)	(0.210)	(0.167)	(0.197)
Nb. Observations	334.00	334.00	334.00	334	334	334	334
R^2	0.18	0.19	0.20	0.18	0.21	0.17	0.13
Weak IV test p.value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Wu-Hausman test p.value	0.150	0.174	0.090	0.132	0.467	0.106	0.053
overindentification testp.value	0.443	0.422	0.512	0.377	0.241	0.370	0.367

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Models 2 and 3 estimated without the largest or the three largest labs

	Without the largest lab		Without the 3 largest labs	
	En. + Ex. effects	En., Ex + FE effects	En. + Ex. effects	En., Ex + FE effects
Intercept	−0.187	–	−0.384	–
	(0.221)		(0.264)	
Age	−0.010**	−0.009*	−0.009	−0.010
	(0.004)	(0.004)	(0.007)	(0.007)
Male	0.117	0.126	0.165*	0.186*
	(0.073)	(0.076)	(0.079)	(0.090)
Distance	0.001	0.002	0.000	0.000
	(0.002)	(0.002)	(0.001)	(0.002)
Yrs in lab	0.007	0.006	0.006	0.008
	(0.006)	(0.006)	(0.009)	(0.011)
Opportunity Cost of active mode (%)	−0.002***	−0.002***	−0.001	−0.001
	(0.000)	(0.000)	(0.000)	(0.001)
Number of children under 12yrs	0.009	0.012	0.024	0.032
	(0.036)	(0.035)	(0.052)	(0.045)
PhD	0.038	0.017	0.073	0.061
	(0.049)	(0.083)	(0.098)	(0.125)
Revenue 4000−6000€	0.008	−0.002	−0.006	−0.024
	(0.091)	(0.081)	(0.136)	(0.130)
Revenue less 2000€	0.058	0.062	−0.031	−0.013
	(0.121)	(0.120)	(0.130)	(0.163)
Revenue more 6000€	0.278**	0.298**	0.246*	0.282
	(0.083)	(0.117)	(0.097)	(0.176)
Health Issue	−0.170	−0.187	−0.198	−0.251
	(0.143)	(0.161)	(0.213)	(0.210)
Environmental Sensibility Index	0.526*	0.579*	0.537	0.653*
	(0.272)	(0.255)	(0.338)	(0.313)
Endogenous Peer Effect	0.598**	0.657	0.406	0.628
	(0.240)	(0.350)	(0.245)	(0.375)
Exogenous Peer Effects	X	X	X	X
Labs' Fixed Effects	–	X	–	X
Nb. Observation	248.00	248.00	162.00	162.00
R^2	0.23	0.56	0.22	0.48
DF residuals	222.00	214.00	136.00	130.00
Weak IV p.value	0.000	0.000	0.000	0.000
Wu-Hausman p.value	0.161	0.087	0.120	0.050
Overindentification p.value	0.520	0.622	0.310	0.236

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Models estimated with G weighted by the intensity of the social relationships

	En. effect only	En. + Ex. effects	En., Ex + FE effects	All effects
Intercept	0.164	−0.200	–	–
	(0.207)	(0.287)	–	–
Age	−0.008	−0.005	−0.004	−0.005
	(0.005)	(0.005)	(0.005)	(0.005)
Male	0.048	0.064	0.062	0.059
	(0.050)	(0.052)	(0.058)	(0.057)
Distance	0.000	0.000	0.000	0.000
	(0.002)	(0.003)	(0.002)	(0.002)
Yrs in lab	0.006	0.004	0.004	0.005
	(0.005)	(0.005)	(0.005)	(0.005)
Opportunity Cost of active mode (%)	−0.001**	−0.001**	−0.002**	−0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Number of children under 12yrs	0.017	0.030	0.032	0.036
	(0.015)	(0.021)	(0.020)	(0.023)
PhD	0.044	0.047	0.016	0.007
	(0.032)	(0.043)	(0.057)	(0.059)
Revenue 4000−6000€	0.033	0.002	−0.002	0.003
	(0.054)	(0.055)	(0.059)	(0.061)
Revenue less 2000€	0.048	0.065	0.036	0.032
	(0.088)	(0.097)	(0.108)	(0.104)
Revenue more 6000€	0.200*	0.187*	0.196	0.198
	(0.093)	(0.080)	(0.117)	(0.122)
Health Issue	−0.167	−0.108	−0.174	−0.180
	(0.095)	(0.070)	(0.100)	(0.095)
Environmental Sensibility Index	0.305	0.384	0.383*	0.395*
	(0.166)	(0.199)	(0.172)	(0.175)
Endogenous Peer Effect	0.481***	0.272	0.562	0.564
	(0.109)	(0.345)	(0.336)	(0.320)
Exogenous Peer Effects	–	X	X	X
Labs' Fixed Effects	–	–	X	X
Gregariousness	–	–	–	0.034
				(0.032)
G_Gregariousness	–	–	–	0.011
				(0.059)
Nb. Observations†	312	312	312	312
R^2	0.20	0.24	0.18	0.18
Weak IV p.value	0.000	0.000	0.000	0.000
Wu-Hausman p.value	0.095	0.878	0.191	0.155
Overindentification p.value	0.286	0.272	0.351	0.331

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

† Some individuals did not answer the questions used to construct the weighted social interaction matrix

Table A.4: Models accounting for heterogeneity in peer effects depending on the commuting distance

	En. effect only	En. + Ex. effects	En., Ex. + FE effects	All effects
Intercept	0.042	0.014	–	–
	(0.163)	(0.318)		
Age	−0.005	−0.002	−0.001	−0.001
	(0.004)	(0.004)	(0.004)	(0.003)
Male	0.098**	0.139*	0.144*	0.141
	(0.040)	(0.070)	(0.073)	(0.074)
Distance	−0.001	0.000	0.000	0.000
	(0.003)	(0.002)	(0.002)	(0.002)
Yrs in lab	0.002	0.001	0.000	−0.001
	(0.005)	(0.005)	(0.005)	(0.004)
Opportunity Cost of active mode (%)	−0.001**	−0.001**	−0.001***	−0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Number of children under 12yrs	0.006	0.044	0.044	0.040
	(0.022)	(0.023)	(0.028)	(0.021)
PhD	0.063	0.091	0.095	0.068
	(0.037)	(0.087)	(0.072)	(0.076)
Revenue 4000−6000€	0.062	−0.020	0.003	0.017
	(0.063)	(0.072)	(0.090)	(0.070)
Revenue less 2000€	0.086	0.052	0.070	0.051
	(0.102)	(0.103)	(0.118)	(0.121)
Revenue more 6000€	0.227**	0.182*	0.178	0.177
	(0.074)	(0.074)	(0.097)	(0.100)
Health Issue	−0.118	−0.235**	−0.207**	−0.188*
	(0.080)	(0.072)	(0.074)	(0.073)
Environmental Sensibility Index	0.437**	0.466	0.482	0.508
	(0.163)	(0.280)	(0.277)	(0.294)
Gyy Peer Effect	0.260	0.601	0.567	0.474
	(0.169)	(0.426)	(0.503)	(0.548)
Gnn Peer Effect	0.042	0.003	0.133	0.167
	(0.294)	(0.417)	(0.500)	(0.492)
Gyn Peer Effect	0.220	0.372	0.307	0.385
	(0.256)	(0.512)	(0.697)	(0.327)
Gny Peer Effect	0.117	0.624	0.162	0.274
	(0.230)	(0.523)	(0.467)	(0.476)
Exogenous Peer Effects	–	X	X	X
Labs' Fixed Effects	–	–	X	X
Gregariousness	–	–	–	0.026
				(0.040)
Nb. Observations	334	334	334	334

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Models accounting for heterogeneity in peer effects depending on having children under 12 years old or not

	En. effect only	En. + Ex. effects	En., Ex. + FE effects	All effects
Intercept	0.146	−0.139	–	–
	(0.243)	(0.226)		
Age	−0.005	−0.001	−0.001	−0.002
	(0.004)	(0.005)	(0.007)	(0.007)
Male	0.128*	0.100	0.097	0.097
	(0.056)	(0.061)	(0.062)	(0.068)
Distance	−0.001	−0.002	−0.002	−0.002
	(0.003)	(0.003)	(0.003)	(0.002)
Yrs in lab	−0.001	0.000	0.002	0.004
	(0.005)	(0.004)	(0.006)	(0.007)
Opportunity Cost of active mode (%)	−0.002**	−0.001*	−0.001*	−0.001**
	(0.001)	(0.001)	(0.001)	(0.000)
Number of children under 12yrs	0.048	−0.029	−0.019	−0.026
	(0.132)	(0.070)	(0.064)	(0.087)
PhD	0.056	0.119	0.084	0.079
	(0.059)	(0.066)	(0.060)	(0.049)
Revenue 4000−6000€	0.075	−0.077	−0.073	−0.056
	(0.072)	(0.057)	(0.056)	(0.052)
Revenue less 2000€	0.114	0.119	0.124	0.119
	(0.101)	(0.102)	(0.094)	(0.093)
Revenue more 6000€	0.235*	0.159**	0.199*	0.216*
	(0.106)	(0.057)	(0.090)	(0.087)
Health Issue	−0.042	−0.225	−0.264*	−0.326**
	(0.164)	(0.122)	(0.100)	(0.104)
Environmental Sensibility Index	0.504*	0.557**	0.648**	0.638**
	(0.227)	(0.200)	(0.188)	(0.195)
Gyy Peer Effect	0.087	−0.241	−0.071	−0.207
	(0.173)	(0.277)	(0.340)	(0.463)
Gnn Peer Effect	−0.094	0.067	0.319	0.439
	(0.129)	(0.566)	(0.445)	(0.320)
Gyn Peer Effect	−0.245	0.416	0.258	0.559
	(0.639)	(0.374)	(0.465)	(0.394)
Gny Peer Effect	0.175	0.377	0.728	0.677
	(0.414)	(0.697)	(0.576)	(0.642)
Labs' Fixed Effects	–	–	X	X
Gregariousness	–	–	–	0.029
				(0.048)
Nb. Observations	334	334	334	334

* p < 0.1, ** p < 0.05, *** p < 0.01

Table A.6: Models of peer effects on the sub-sample complete network observations

	En. effect only	En. + Ex. effects	En., Ex. + FE effects	All effects
Intercept	0.327	0.404	–	–
	(0.316)	(0.576)		
Age	−0.010	−0.013	−0.015	−0.006
	(0.007)	(0.009)	(0.007)	(0.011)
Male	0.041	0.041	−0.028	−0.043
	(0.183)	(0.143)	(0.189)	(0.131)
Distance	−0.002	−0.006	−0.004	−0.007
	(0.004)	(0.007)	(0.004)	(0.005)
Yrs in lab	−0.001	0.001	0.002	−0.007
	(0.005)	(0.005)	(0.005)	(0.006)
Opportunity Cost of active mode (%)	−0.001	0.000	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Number of children under 12yrs	−0.041	−0.067	−0.025	−0.019
	(0.068)	(0.078)	(0.069)	(0.071)
PhD	0.252*	0.324**	0.284*	0.466**
	(0.104)	(0.104)	(0.124)	(0.136)
Revenue 4000−6000€	0.168	0.127	0.121	−0.007
	(0.106)	(0.113)	(0.063)	(0.143)
Revenue less 2000€	0.048	0.031	0.001	0.155
	(0.106)	(0.153)	(0.082)	(0.123)
Revenue more 6000€	0.232	0.189	0.205	0.573*
	(0.218)	(0.272)	(0.226)	(0.235)
Health Issue	−0.067	−0.142	−0.141	0.422
	(0.303)	(0.249)	(0.384)	(0.256)
Environmental Sensibility Index	0.318	0.436	0.193	0.240
	(0.280)	(0.437)	(0.201)	(0.315)
Endogenous Peer Effect	0.362	0.491	0.420	0.724
	(0.285)	(0.469)	(0.363)	(0.415)
Exogenous Peer Effects	–	X	X	X
Labs' Fixed Effects	–	–	X	X
Gregariousness	–	–	–	−0.116
				(0.096)
G_Gregariousness	–	–	–	0.039
				(0.174)
Nb. Observation	72.00	72.00	72.00	72.00
R^2	0.29	0.43	0.29	0.55
Weak IV p.value	0.000	0.000	0.000	0.000
Wu-Hausman p.value	0.914	0.991	0.515	0.854
Sargan p.value	0.440	0.421	0.287	0.091

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Models of peer effects on Sustainable Transportation's choice

	En. effect only	En. + Ex. effects	En., Ex. + FE effects	All effects
Intercept	0.612*	0.484	–	–
	(0.308)	(0.335)		
Age	−0.011**	−0.010*	−0.010*	−0.010*
	(0.004)	(0.005)	(0.004)	(0.004)
Male	−0.010	0.007	0.008	0.006
	(0.047)	(0.049)	(0.051)	(0.048)
Distance	−0.002	−0.002	−0.002	−0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Yrs in lab	0.006	0.007*	0.006	0.006*
	(0.003)	(0.003)	(0.003)	(0.003)
Opportunity Cost of sustainable mode (%)	−0.001	−0.001	−0.001	−0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Number of children under 12yrs	−0.051	−0.044	−0.040	−0.042
	(0.032)	(0.032)	(0.028)	(0.030)
PhD	0.093	0.113	0.129*	0.111
	(0.050)	(0.059)	(0.062)	(0.065)
Revenue 4000−6000€	−0.025	−0.050	−0.049	−0.034
	(0.084)	(0.080)	(0.081)	(0.091)
Revenue less 2000€	−0.043	−0.045	−0.023	−0.026
	(0.054)	(0.053)	(0.045)	(0.046)
Revenue more 6000€	0.037	0.010	0.012	0.016
	(0.115)	(0.117)	(0.112)	(0.118)
Health Issue	−0.009	0.001	0.001	−0.027
	(0.102)	(0.115)	(0.110)	(0.119)
Environmental Sensibility Index	0.209	0.257	0.241	0.229
	(0.176)	(0.189)	(0.179)	(0.178)
Endogenous Peer Effect	0.527*	0.444	0.324	0.350
	(0.217)	(0.266)	(0.300)	(0.281)
Exogenous Peer Effects	–	X	X	X
Labs' Fixed Effects	–	–	X	X
Gregariousness	–	–	–	0.034
				(0.040)
G_Gregariousness	–	–	–	−0.062
				(0.035)
Nb. Observation	334	334	334	334
R^2	0.11	0.16	0.17	0.18
Weak IV p.value	0.000	0.000	0.000	0.000
Wu-Hausman p.value	0.149	0.430	0.425	0.446
overindentification p.value	0.711	0.422	0.513	0.517

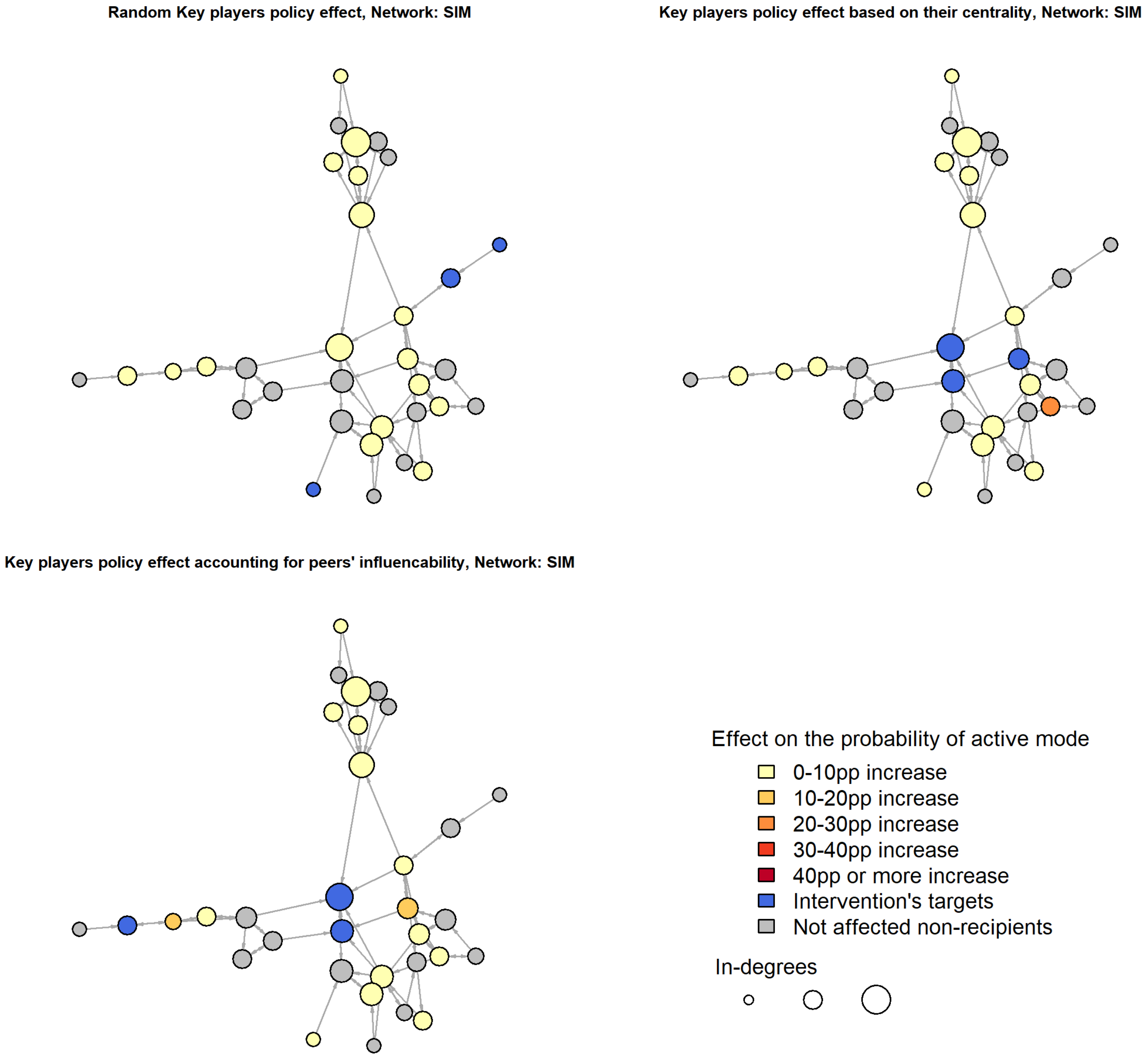
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Estimation of the complete model of peer effects with a Simulated Maximum Likelihood à la Lee et al (2014)

	Estimate	Std.Error	P.value	Marginal Effect
Intercept	-1.264***	0.308	0.000	-0.852
Age	-0.014**	0.006	0.018	-0.009
Male	0.215***	0.062	0.001	0.145
Distance	-0.036***	0.003	0.000	-0.024
Yrs in lab	0.005	0.005	0.330	0.004
Opportunity Cost of active mode (%)	-0.005***	0.001	0.000	-0.003
Number of children under 12yrs	0.107***	0.038	0.005	0.072
PhD	0.141*	0.081	0.082	0.095
Revenue 4000−6000€	0.032	0.080	0.693	0.021
Revenue less 2000€	-0.006	0.106	0.952	-0.004
Revenue more 6000€	0.663***	0.114	0.000	0.431
Environmental Sensibility Index	1.148***	0.180	0.000	0.773
Health Issue	-0.346***	0.123	0.005	-0.229
Endogenous Peer Effect	0.282**	0.139	0.042	0.190
G_Age	0.011	0.011	0.322	0.007
G_Male	0.312***	0.118	0.008	0.210
G_Distance	-0.019**	0.008	0.026	-0.013
G_Yrs in lab**	-0.008	0.012	0.480	-0.006
G_Opportunity Cost of active mode (%)**	0.001	0.001	0.713	0.000
G_Number of children under 12yrs	0.216**	0.067	0.001	0.145
G_PhD	-0.013	0.175	0.942	-0.009
G_Revenue 4000−6000€	0.042	0.133	0.752	0.028
G_Revenue less 2000€	0.377	0.257	0.143	0.254
G_Revenue more 6000€	0.347*	0.204	0.089	0.233
G_Environmental Sensibility Index	-0.034	0.331	0.919	-0.023
G_Health Issue	1.303***	0.201	0.000	0.877
Labs' Fixed Effects	X	–	–	–
Gregariousness	0.153***	0.035	0.000	0.103
G_Gregariousness	0.094	0.076	0.217	0.063
Nb. Observation	334	334	334	334

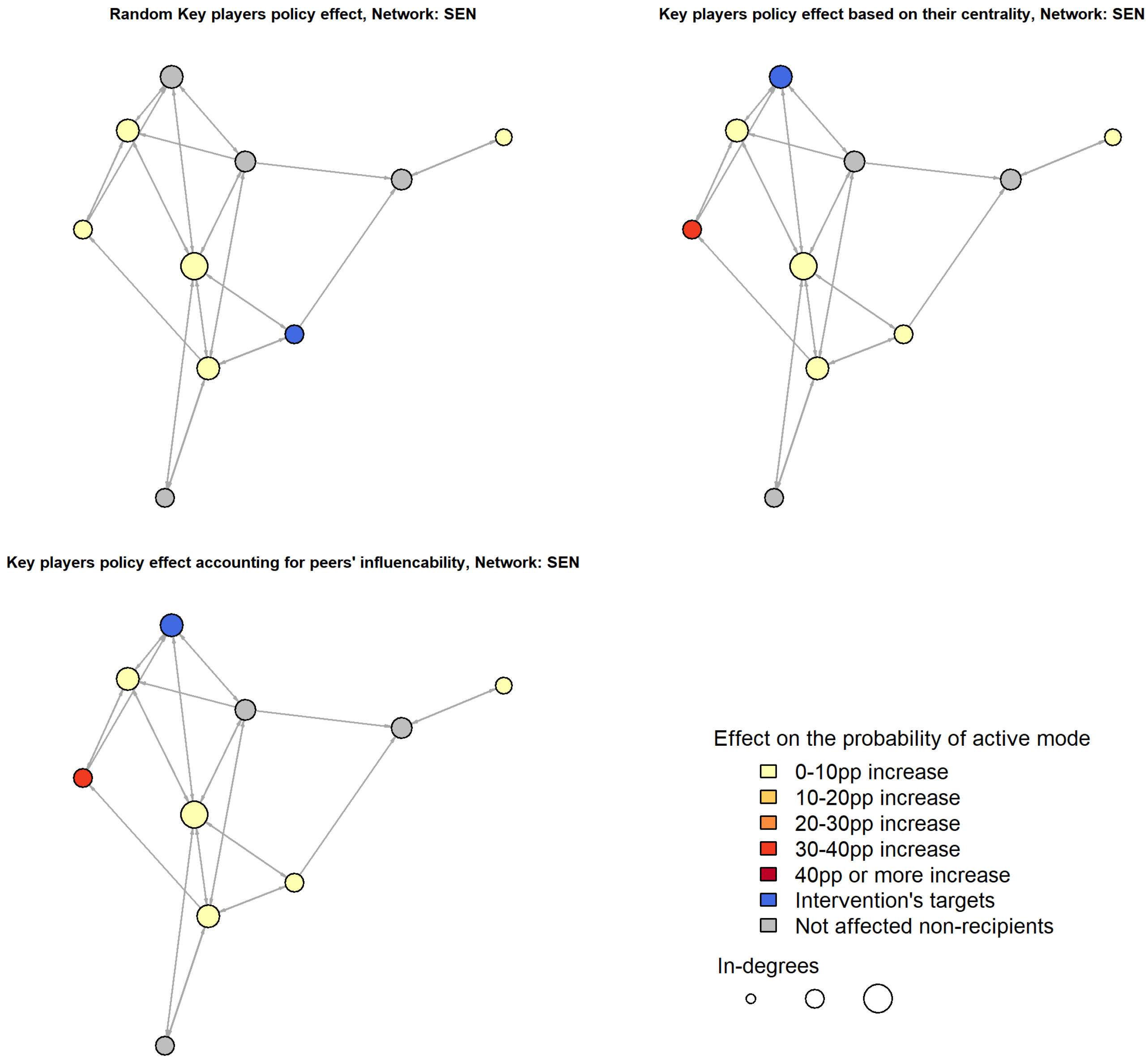
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure A.2: Comparison of spillovers with different selection strategies, SIM



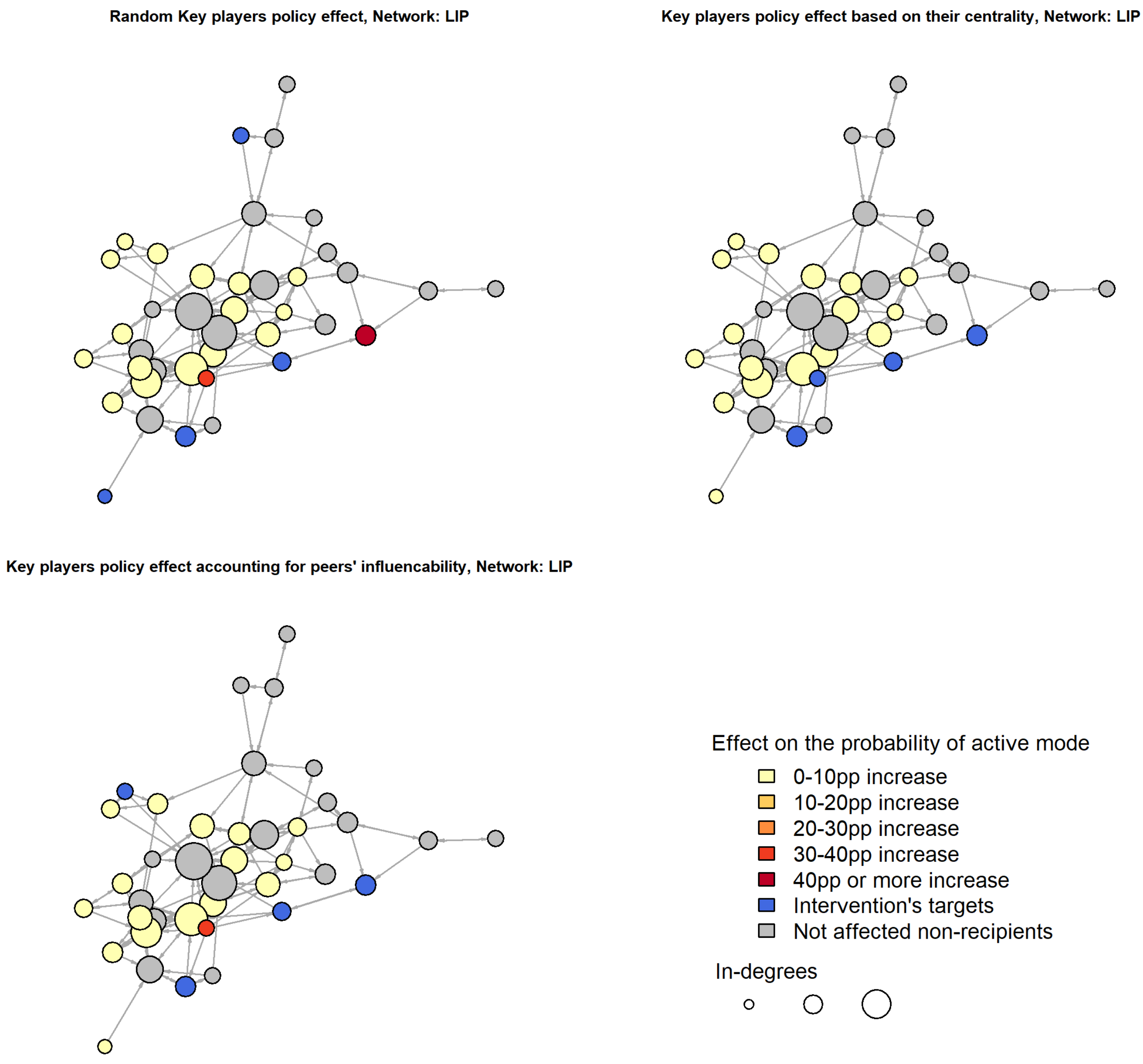
Note: Not affected non-recipients of the intervention are either already using an active transportation mode or living 15km or more away from their workplace.

Figure A.3: Comparison of spillovers with different selection strategies, SEN



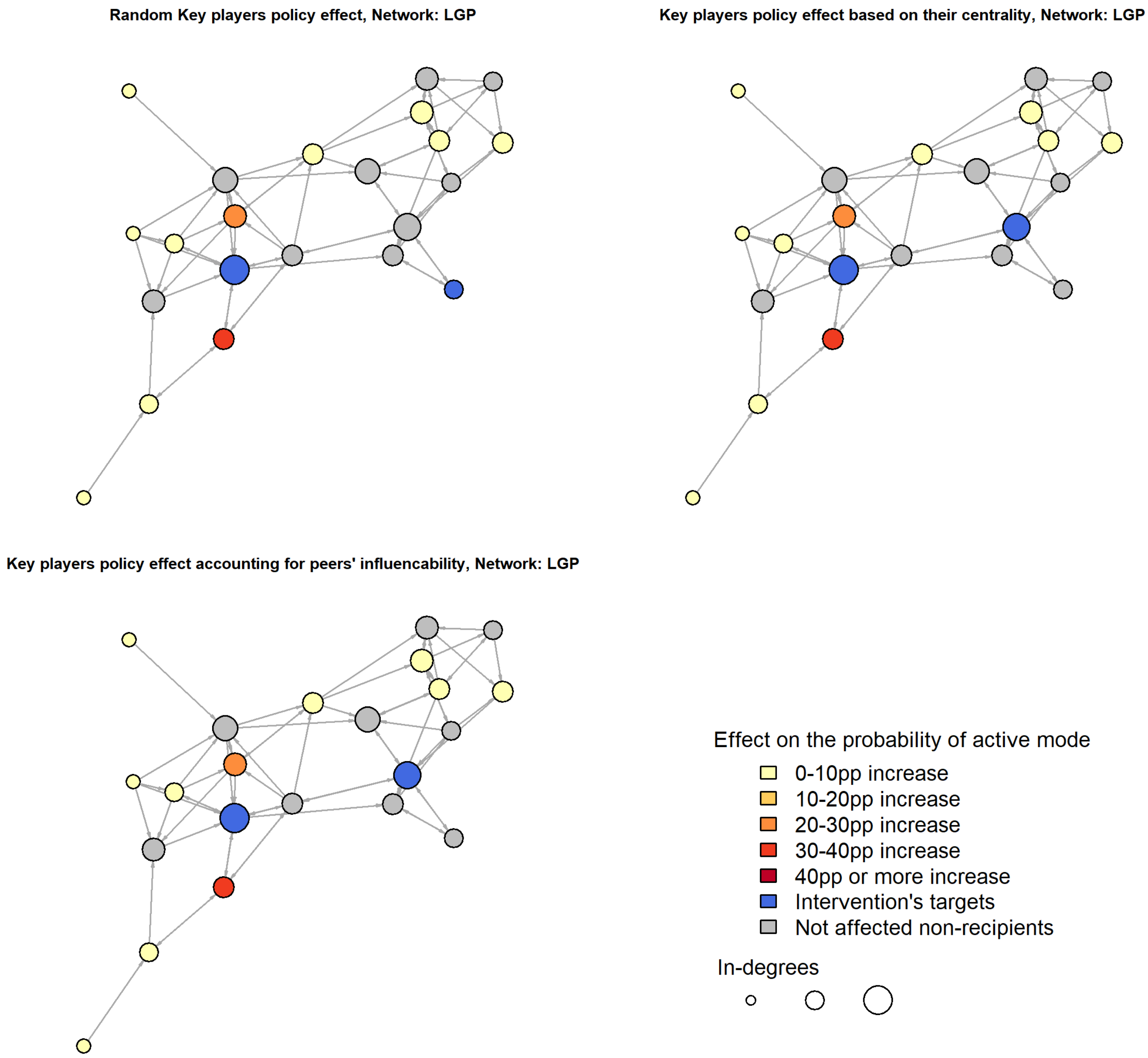
Note: Not affected non-recipients of the intervention are either already using an active transportation mode or living 15km or more away from their workplace.

Figure A.4: Comparison of spillovers with different selection strategies, LIP



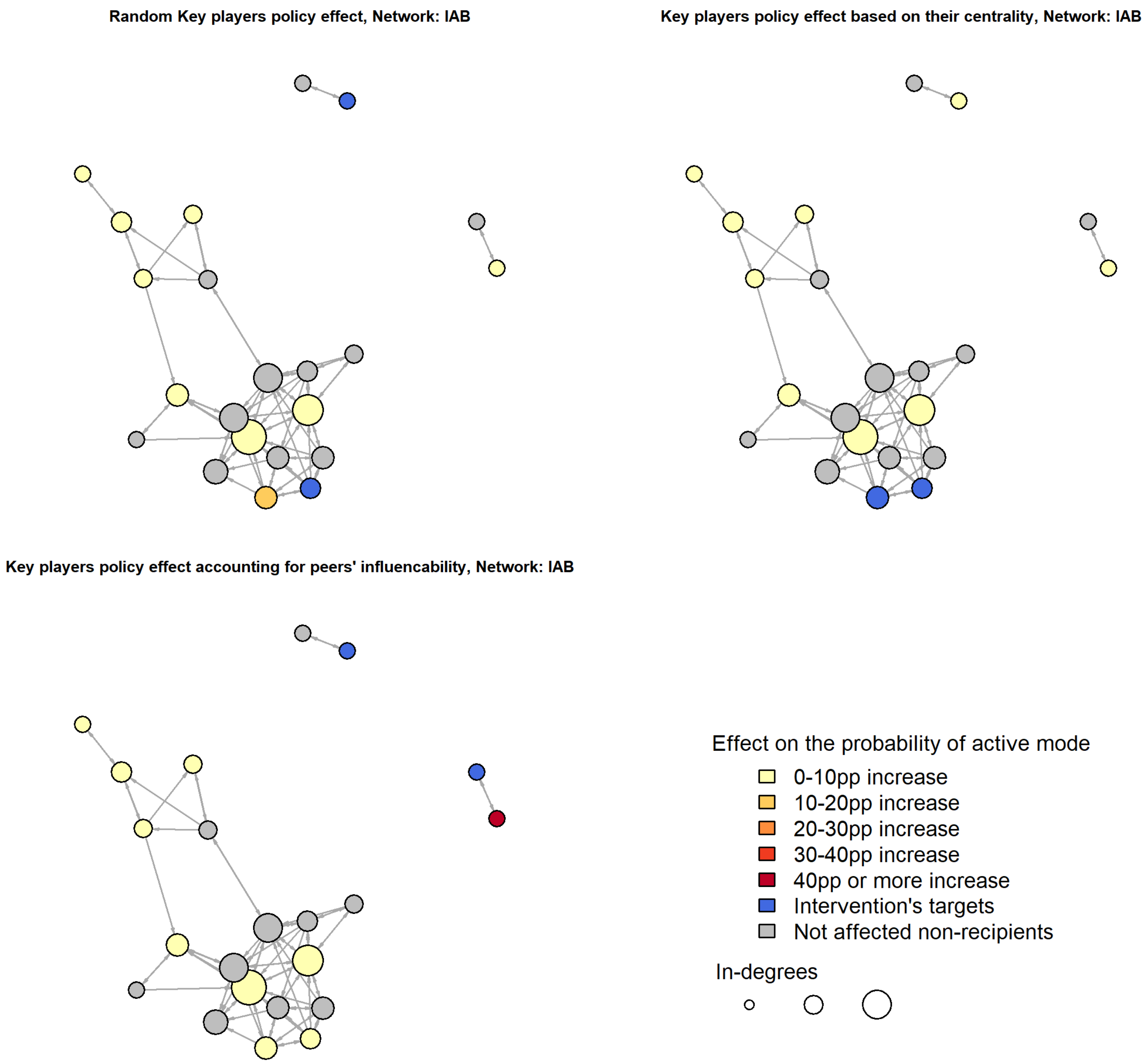
Note: Not affected non-recipients of the intervention are either already using an active transportation mode or living 15km or more away from their workplace.

Figure A.5: Comparison of spillovers with different selection strategies, LGP



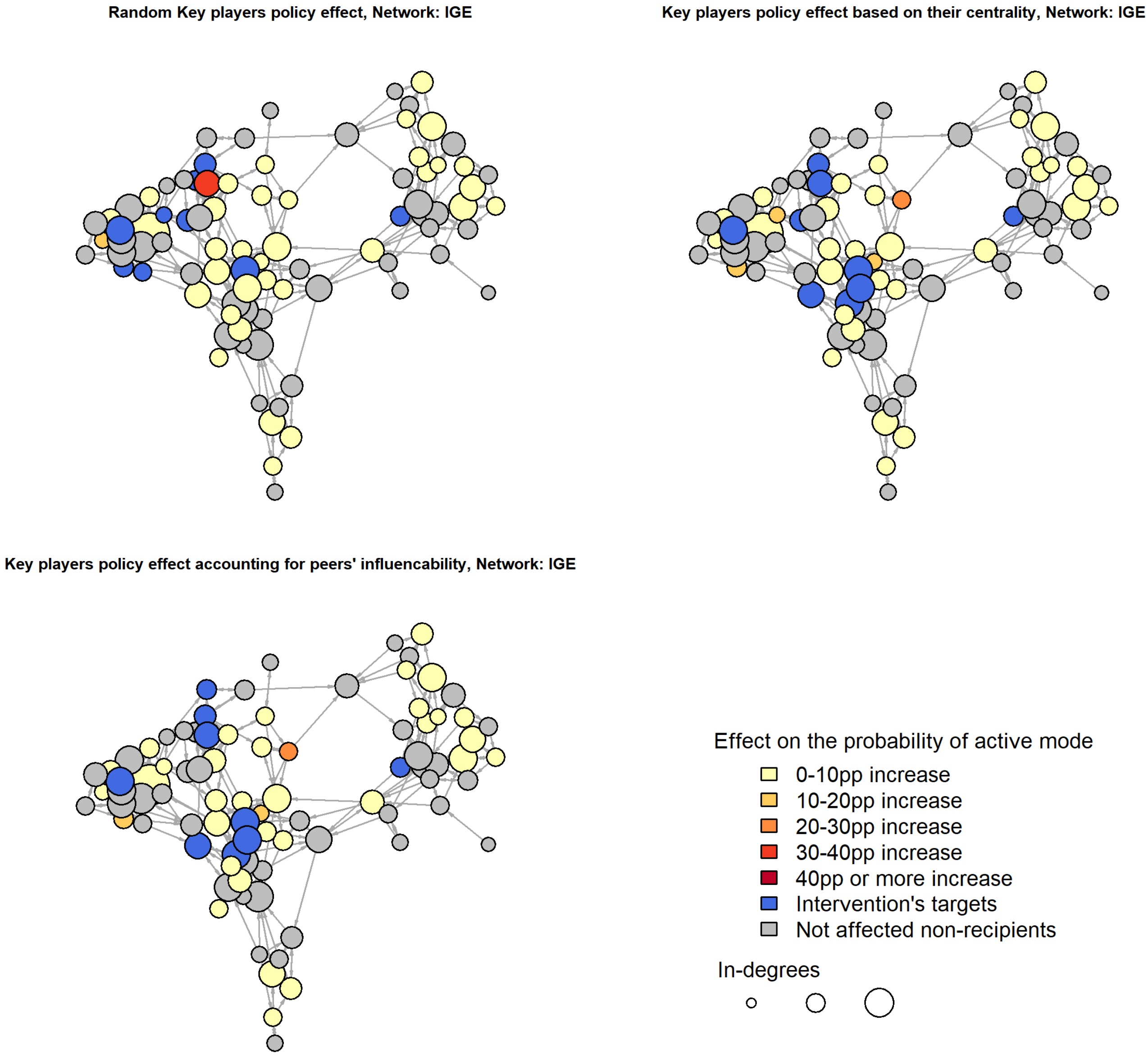
Note: Not affected non-recipients of the intervention are either already using an active transportation mode or living 15km or more away from their workplace.

Figure A.6: Comparison of spillovers with different selection strategies, IAB



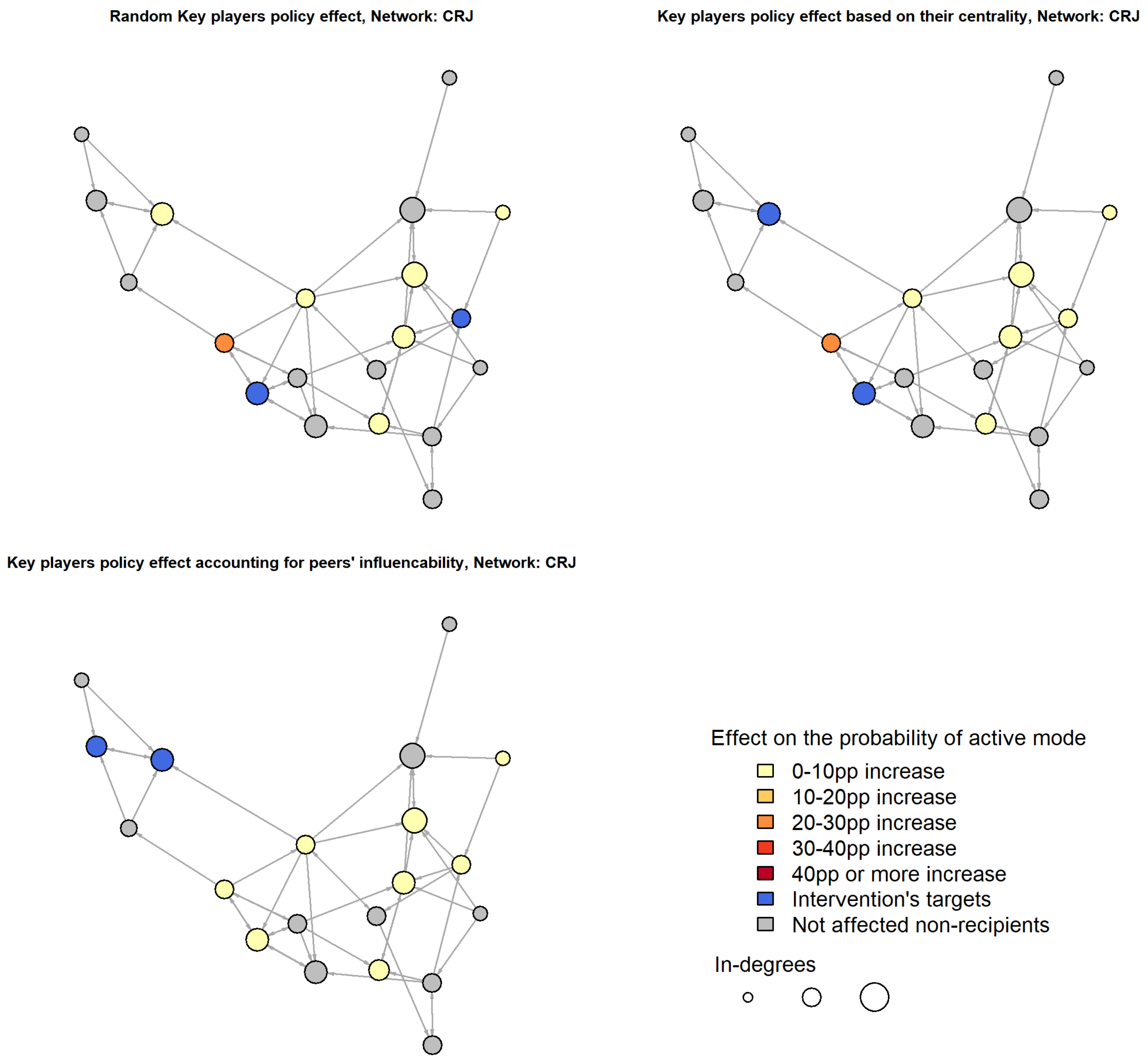
Note: Not affected non-recipients of the intervention are either already using an active transportation mode or living 15km or more away from their workplace.

Figure A.7: Comparison of spillovers with different selection strategies, IGE



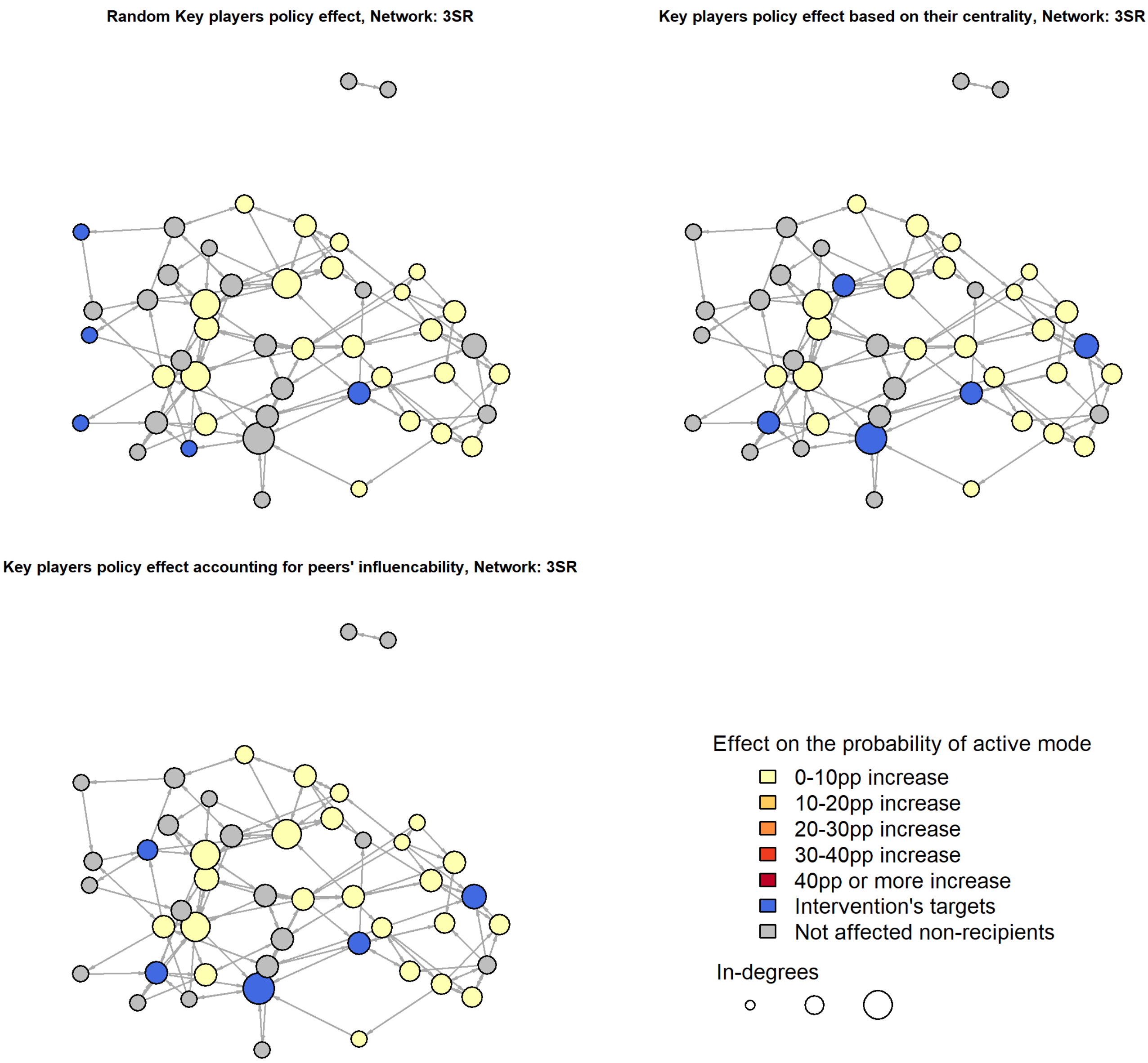
Note: Not affected non-recipients of the intervention are either already using an active transportation mode or living 15km or more away from their workplace.

Figure A.8: Comparison of spillovers with different selection strategies, CRJ



Note: Not affected non-recipients of the intervention are either already using an active transportation mode or living 15km or more away from their workplace.

Figure A.9: Comparison of spillovers with different selection strategies, 3SR



Note: Not affected non-recipients of the intervention are either already using an active transportation mode or living 15km or more away from their workplace.