Externalities of Marijuana Legalization: Marijuana Use in Non-Legalizing States*

Marit Hinnosaar† Elaine M. Liu‡ Eva Loaeza-Albino§

September 13, 2024

Abstract

We study the impact of distant connections on marijuana use. Using data on Facebook friendships, we investigate whether connections to states where recreational marijuana is legalized affect marijuana use in areas where it remains illegal. We find that areas more connected to legalized states have a larger increase in marijuana use. These results indicate that even non-local connections can significantly impact behavior. Our findings suggest that studies measuring the impact of marijuana legalization using a standard difference-in-differences approach, without accounting for the network, underestimate the direct effect on the state that legalized.

Keywords: marijuana, externalities, peer effect, Facebook

JEL Codes: I12, I18

*We thank Sonia Bhalotra, Clément de Chaisemartin, Laura Dague, Toomas Hinnosaar, Rosalie Pacula, Maria Petrova, and conference participants at ASHEcon, Texas STATA Empirical Micro, SEA, SOLE, Princeton Industrial Relations Section Centenary Symposium, and seminar participants at Norwegian School of Economics, University of Zurich, University of Navarra, University of Exeter, University of Autonoma-Madrid, University of Gothenburg, University of Autonoma-Barcelona, University of Copenhagen, and Pompeu Fabra University for helpful comments and suggestions. We also thank Riley Wilson and Michael Bailey for generously sharing the data.

†University of Nottingham and CEPR, marit.hinnosaar@gmail.com.

‡University of Houston, NBER, IZA, and HCEO, emliu@uh.edu.

§California Workers' Compensation Institute, eva.loaeza.albino@gmail.com.

1 Introduction

Many studies have demonstrated the influence of peers on individuals' behaviors and outcomes (Sacerdote, 2001; Kremer and Levy, 2008; Carrell, Fullerton and West, 2009; Sacerdote, 2011; Card and Giuliano, 2013). Typically, these studies focused on how peers physically close to someone, like roommates or classmates, affect them. However, the landscape of peer dynamics has undergone a dramatic transformation in the past two decades due to the widespread adoption of smartphones and the pervasiveness of social media. In the era preceding the advent of smartphones, interactions with acquaintances were only sporadic throughout the year and held limited sway over one's decisions and behaviors. The rise of smartphones has significantly altered this dynamic, enabling geographically distant peers to exert their influence via messaging apps and social media platforms. In the context of this paper, we aim to examine the extent to which these geographically dispersed peers can shape our behaviors.

Nevertheless, similar to research on close-proximity peers, identifying peer effects poses a challenge due to the endogenous selection of peers. Individuals who share similarities tend to naturally gravitate toward each other as peers, making it difficult to distinguish between peer effects and the influence of other factors. In our analysis, we employ a strategy akin to the one proposed by Wilson (2022). We operate under the assumption that social networks remain fixed and leverage temporal variations in the enactment of state-level marijuana legalization laws. First, we explore changes in one's own marijuana use resulting from marijuana legalization changes within their own state. Subsequently, we examine whether such changes extend to the marijuana use patterns of distant peers in an aggregated fashion, which will be described in more detail later.

The study of marijuana use presents an interesting case due to its varied legal status across the United States. Marijuana use remains illegal federally; however, as of July 2024, 24 states and Washington D.C. have legalized recreational marijuana. Remarkably, approximately 20% of American adults aged 18 and older reported marijuana use in the past year. This figure surges to 38% among individuals aged 21 to 25, and even within the 41-45 age group, 21% acknowledged marijuana use within the past year.[1] Many of these individuals reside in states where marijuana use remains illegal under federal and state law. Previous research on marijuana legalization has indicated that state-level legalization leads to increased marijuana use within the legalized states (Hollingsworth, Wing and Bradford, 2022). Additionally, other research has explored the cross-border spillover effects of marijuana legalization (Hansen,

[1]Source: 2021 NSDUH Detailed Tables, Table 1.7B, published January 4, 2023, accessed August 31, 2023, https://www.samhsa.gov/data/sites/default/files/reports/rpt39441/NSDUHDetailedTabs2021/NSDUHDetailedTabs2021/NSDUHDetailedTabs2021.htm#tab1.7b

Miller and Weber, 2020). Our study delves into the intriguing question of how one state's marijuana legalization can impact even geographically distant areas, extending beyond the immediate bordering localities.

Our analysis is conducted at the aggregated level, utilizing the Social Connectedness Index (SCI) created by Facebook as our measure of social connections. The SCI reflects the likelihood of Facebook connections between residents of any two US counties and any two zip codes. While Facebook users do not represent a fully representative sample of the population, a PEW Research Center Survey (2021) reported that 69% of American adults have used Facebook at some point, and this index has been employed in other economic research papers (including Bailey et al. (2018*c*, 2019); Wilson (2022)). Furthermore, we use this index as a proxy for general social connections. To the extent that the SCI constructed from Facebook connections may introduce noise in measuring overall social connections, our analysis might be conservative, providing a lower estimate of the true impact. Using this index, we construct a measure of exposure to legalized marijuana, which varies across localities and time based on the strength of connections to legalized states and the timing of legalization. We gauge marijuana consumption using two aggregate-level datasets: self-reported marijuana usage data from the National Survey of Drug Use and Health (NSDUH) and workplace marijuana testing positivity rates provided by Quest Analytics. One concern with the NSDUH is that marijuana legalization might affect respondents' honesty in reporting marijuana use. While the Quest dataset is not a representative sample of the population, it helps alleviate concerns related to the potential divergence between self-reported marijuana usage and actual usage.

We regress self-reported marijuana use and workplace marijuana positivity rates on an imputed measure of exposure to legalized marijuana, incorporating both locality fixed effects and state-year pair fixed effects. By including state-year pair fixed effects, we compare localities with different levels of exposure within the same state in a given year. Our results show that both self-reported marijuana use and workplace positivity rates increase more in areas more closely connected to states where recreational marijuana is legal. Specifically, a one standard deviation increase in exposure to these states corresponds to a roughly one-quarter standard deviation increase in self-reported use and a one-sixth standard deviation increase in workplace positivity rates.

Can we attribute all these effects to spillover from Facebook or online social interactions? The short answer is no. Facebook friends are often also friends in real life, and they may communicate in person or via other platforms. Thus, we cannot attribute the entire impact to Facebook "likes" and comments. However, when we examine the heterogeneity of the exposure impact by physical distance, we find that the influence of distant peers (greater than 190 miles) is as strong as that of nearby peers. This insight highlights that in today's interconnected world, geographical distance matters less, as individuals can still exert substantial influence on one another regardless of proximity.

One threat to our identification is that regions with higher exposure to legalized marijuana might have different trends in marijuana use compared to regions with lower exposure. Specifically, high-exposure regions could have experienced an increase in marijuana use regardless of exposure. To address this potential issue, we conducted several robustness checks. For example, we included a set of baseline characteristics correlated with the measure of marijuana legalization exposure and allowed these characteristics to have differential impacts on marijuana smoking. Additionally, since our marijuana smoking outcomes are only measured until 2018 and 2020, we also use legalization exposure resulting from future law changes (occurring between 2020 and 2023) as a control in our regressions. Having this future exposure control should allow us to control, to some extent, the underlying difference between localities. Our results remained consistent and robust across these checks.

Contribution to the literature. Our study adds to the research on spatial externalities of policies, in particular, the cross-border effects of substance use regulations. Prior studies have mainly explored cross-border effects for alcohol (Lovenheim and Slemrod, 2010; Johansson, Pekkarinen and Verho, 2014; Jacks, Pendakur and Shigeoka, 2021) and cigarettes (Merriman, 2010; Harding, Leibtag and Lovenheim, 2012; DeCicca, Kenkel and Liu, 2013). In the context of marijuana, Hansen, Miller and Weber (2020) analyzed the impact of cross-border shopping on sellers' profits. The above research has used geographic proximity, typically neighboring status, to quantify spatial externalities. Our work focuses on a different measure of connectedness. The measure we are using can include strong connections between even distant regions. We show that the measure facilitates the spread of marijuana use beyond the borders of the state that legalized. While we don't rule out that the impact of Facebook connectedness is driven by some form of geographic connections, we believe Facebook connectedness provides another approach to model spatial externalities in this setting.

In addition, our results highlight the importance of spatial externalities of policies. If legalization in one state has externalities to other more connected out-of-state areas, then studies that estimate the impact of legalization using a standard difference-in-differences approach comparing states that legalized to those that did not, without taking into account the externalities, underestimate the direct effect of legalization on the state that legalizes.

Our paper also contributes to the literature on the impact of marijuana legalization on marijuana consumption (including Williams and Bretteville-Jensen (2014); Jacobi and Sovinsky (2016).[2] Hollingsworth, Wing and Bradford (2022) have analyzed the effects of recent changes, including recreational marijuana legalization within the United States, focusing on the impact within the same state. In contrast, our paper examines the externalities in areas beyond the legalized states, broadening the scope of analysis.

In the literature on peer effects in risky behaviors, existing economic studies predominantly rely on a quasi-random exposure to fellow students within the same school (Lundborg, 2006; Clark and Lohéac, 2007) or random roommate assignments (Eisenberg, Golberstein and Whitlock, 2014). Our approach differs from previous studies in that we consider networks as given, albeit potentially endogenous, and utilize a policy shock for identification to assess changes in behavior within existing social networks.

Finally, our paper adds to the recent list of papers that study spillovers using the Facebook Social Connectedness Index (including Bailey et al. (2018*c*, 2019); Wilson (2022)). The closest to our work are papers by Cutler and Donahoe (2024) and Mäckle and Ruenzi (2023) that document large spillovers in opioid use. We examine the impact of Facebook connections on another risky behavior, specifically marijuana smoking. Marijuana use presents a compelling case for analysis for two primary reasons. Firstly, its legality varies depending on one's state of residence. Our findings reveal that when friends' states legalize marijuana, it influences the marijuana consumption of individuals residing in areas where marijuana is still illegal. Secondly, compared to opioids, marijuana use stands out as a more prevalent social activity. Given its social nature, we might anticipate that the peer effects on marijuana smoking are more pronounced compared to the spillovers estimated in other settings.[3] On the other hand, compared to opioids, marijuana is less addictive, which might lead to smaller spillovers. While we estimate sizable spillovers, the effects on marijuana use are smaller than those documented in Cutler and Donahoe (2024) for opioids.

The paper is organized as follows. Section 2 provides background on marijuana policy. Sections 3 and 4 describe the data and the empirical methodology. Section 5 reports the results. Section 6 concludes.

[2]Literature on marijuana legalization has also studied effects on consumption of other substances including alcohol, tobacco, opioids, prescription painkillers, and illegal drugs (Kelly and Rasul, 2014; Wen, Hockenberry and Cummings, 2015; Bradford and Bradford, 2018; Powell, Pacula and Jacobson, 2018; Carrieri, Madio and Principe, 2020; McMichael, Van Horn and Viscusi, 2020; Sovinsky et al., 2023), traffic fatalities (Anderson, Hansen and Rees, 2013), crime (Adda, McConnell and Rasul, 2014; Gavrilova, Kamada and Zoutman, 2019; Dragone et al., 2019), education (Marie and Zölitz, 2017). Other topics the literature has analyzed include the regulatory policy of the legal marijuana market (Hollenbeck and Uetake, 2021) and cross-border trade (Hansen, Miller and Weber, 2020).

[3]Vannucci et al. (2020) provides an overview of papers that examine the relationship between social media use and risky behaviors. All of these papers point to a correlation between the use of social media and risky behaviors.

2 Marijuana legalization

In the US, the legalization of recreational marijuana followed the earlier legalization of medical marijuana, which began as early as 1996 in California. Our analysis specifically focuses on the legalization of recreational marijuana, which began in Washington and Colorado in 2012. As of July 2024, recreational marijuana has been legalized in 24 states and Washington, DC.[4] For a comprehensive list of recreational marijuana legalization events, please refer to Table A.1 and Figure A.1 in the Online Appendix.

The legalization of recreational marijuana can be viewed as a two-step process. The first step involves legalizing the use of marijuana, while the second step entails licensing the sale of marijuana. The latter step ensures that recreational marijuana is readily available to consumers. In our primary analysis, we will focus on the impact of marijuana use legalization. The analysis of licensed sales is presented as a robustness check (Figure 3) and the results are similar to our main estimates.

3 Data

The analysis utilizes three main data sources. To measure marijuana use, we utilize the National Survey on Drug Use and Health (NSDUH) and the Drug Testing Index from Quest Diagnostic. To measure connectedness, we use Facebook's Social Connectedness Index (SCI).

National Survey on Drug Use and Health. NSDUH is a nationally representative survey of the US population aged 12 and above. Specifically, we rely on the NSDUH Substate public data, which provides the prevalence of marijuana use in the previous year and the previous month. Previous studies in the marijuana literature have also used the NSDUH aggregated public data (Hollingsworth, Wing and Bradford, 2022; Choi, Dave and Sabia, 2019). The NSDUH Substate report combines data from multiple years (2002-2004, 2004-2006, 2006-2008, 2008-2010, 2010-2012, 2012-2014, 2014-2016, 2016-2018).[5] In the paper, to simplify the exposition, we refer to each 3-year interval by the interval end year. We restrict the sample to 245 substates (out of 406 substates) where boundaries have remained constant over time, representing approximately 80% of the US contiguous population. See Figure A.2 for substate sample coverage. Our main outcome variable is the percentage of populations

[4]Source: `https://en.wikipedia.org/wiki/Legality_of_cannabis_by_U.S._jurisdiction`, accessed July 24, 2024.

[5]The NSDUH substate 2018-2020 report was originally available, but as of July 2023, the NSDUH website has removed the 2018-2020 data citing issues with inconsistent modes of data collection during Covid. See 2016-2018 NSDUH Overview and Summary Substate Region Estimation Methodology.

aged 18 and above that report using marijuana at least once in the past year in a given substate. Later, we also investigate marijuana use in other age groups.

Marijuana Positivity Rates from Quest Diagnostic. To address concerns about self-reporting bias, we supplement our analysis with data from nonsurvey sources. Specifically, we incorporate data from Quest Diagnostics, which provides annual workplace marijuana positivity rates at the 3-digit zip code level from 2007 to 2020. Quest Diagnostics, a commercial laboratory, conducts workplace drug tests on behalf of employers. In recent years (2017-2020), Quest has conducted nearly 8 million tests annually. However, this dataset reflects a non-representative sample of the U.S. workforce, as it includes only employees whose employers use Quest Diagnostics for drug testing. This sample likely differs from the general population, as it primarily includes individuals either recently offered positions or employed in roles requiring regular drug tests. These individuals would have lower rates of marijuana use compared to the general population. Given the sample selection issue of this dataset, we are viewing the coefficients from the analysis conducted with this dataset as the average treatment effect for this particular group, which might differ from the average treatment effect of the entire population. For all the years 2007-2020, the marijuana positivity rate is available for 84% of the zip codes with residential addresses.[6] Our main sample includes 753 3-digit zip codes. The marijuana positivity rate in about three-quarters of the time is provided not as a single number, but instead, as an interval. For our main analysis, we use the lower bound of the interval. In the robustness analysis, we use the median and the upper bound of the interval. To make it more comparable to outcomes from NSDUH dataset, we calculated a 3-year running average of the positivity rate (e.g., 2007-2009, 2010-2012, etc.). In the robustness analysis, we also provide estimates without using the 3-year averaging. This dataset has also been used by several other papers that examine the impact of marijuana legalization (Abouk, Mansouri and Powell, 2024; Dong, 2022; Hollingsworth, Wing and Bradford, 2022).

Facebook's Social Connectedness Index. The SCI is based on friendship links on Facebook and measures the relative probability of two individuals from different locations being friends on Facebook (Bailey et al., 2018*a*).[7] The SCI data, provided to us from Facebook, is available at two levels and two time periods: county-to-county (for both 2016

[6]In the US, there are more than 900 3-digit zip codes in use, but some of these are not residential addresses. We focus on the continental US and exclude from the analysis 3-digit zip codes related to the Washington government, military, and Guam.

[7]It is a relative probability since the probability was scaled from between 0 and 1 to be between 1 and 1,000,000,000.

and 2021) and 5-digit zip code-to-zip code (only available for the year 2021). To match the level of the main outcome variables, we aggregate the SCI based on population weights at the substate level (using county-level SCI) and the 3-digit zip code level (using 5-digit zip code-level SCI). More details about the construction of population-weighted SCI are presented in Appendix C.

The main variable of interest: exposure to legalization. We would like to measure how exposure via the social network to out-of-state areas that have legalized marijuana affects marijuana use. To do that, first, we construct for each locality a measure of exposure as a population-weighted average of the Social Connectedness Index across all the localities where marijuana is legal in a given period i except those in the current state. The network exposure measure in locality i in state s in period t equals:

$$Exposure_{it} = \sum_{j \notin s}^{J} 1[LegalUseIn\ j]_t \times Population_j \times SCI_{ij} \tag{1}$$

where $1[LegalUseIn\ j]_t$ is an indicator for legal use in locality j in period t, $Population_j$ is the total population in locality j, and SCI_{ij} is the Social Connectedness Index between locality i and locality j. The measure captures for each person in locality i the relative number of friends in areas where marijuana use is legal. It is a measure of a relative number of friends instead of an absolute number of friends since SCI_{ij} is a relative probability measure. We construct this network exposure variable at both locality levels: substate level and 3-digit zip code level. 3-digit zip code exposure is constructed using 2021 zip code-zip code SCI while the substate exposure is constructed using 2021 county-county SCI. In the robustness checks, since the 2016 SCI is not available at the zip code level, we have also used the 2016 county-county SCI to construct alternative substate exposure, and the results are robust. In the analysis, we use the logarithm of the exposure measure from the previous literature (Bailey et al., 2020, 2021).

Geographic measures and economic conditions. To investigate the role of physical connectedness in addition to the online social network, we use three sets of measures. First is an indicator variable for whether the locality is neighboring any legalized state. Second is an indicator for any direct flights between the locality and the legalized states. Last, we calculate the minimum distance from each locality to any state that has legalized marijuana using the centroid coordinates of each locality. The minimum distance variable for years prior to 2012 (first state marijuana legalization) was set to a constant of two times the maximum post-2012. For the econometric analysis, the exact specification of the distance prior to 2012 does not matter, as state-year fixed effects would have absorbed it. When a state has legalized marijuana, the minimum distance would be 0. In our analysis, we include a standardized version of this variable as a control.

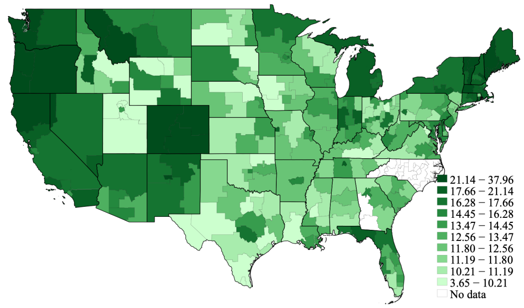
To control for current economic conditions, we use per capita income from the Bureau of Economic Analysis and the unemployment rate from the Bureau of Labor Statistics. We calculate the income to 2018 dollars using the consumer price index for all urban consumers.

Summary statistics. Table 1 provides a snapshot of the key variables in the most recent period available. On average, 14.6 percent of the population aged 18 and older reported using marijuana between 2016 and 2018. Unsurprisingly, workplace marijuana positivity rates were much lower, averaging 2 percent in 2020. Notably, there is considerable variation in both marijuana use and exposure to legalization across localities within many states(Figure 1).

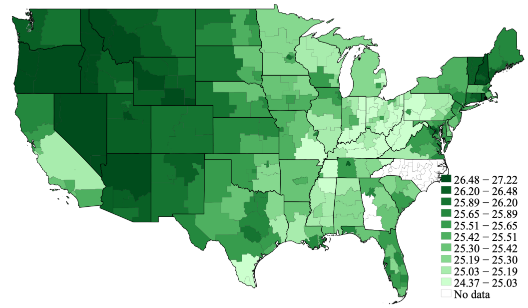
Table 1: Summary statistics

	mean	sd	min	max
Panel A: Substate Level Data				
Marijuana use in the past year	14.55	4.36	6.35	39.37
Log of network exposure	25.62	0.56	24.37	27.22
Standardized min. distance to a legalized state	-0.00	1.00	-1.53	2.31
Neighboring a legalized state	0.13	0.34	0.00	1.00
Having a direct flight to a legalized state	0.91	0.28	0.00	1.00
Unemployment rate	4.33	1.05	2.17	8.23
Per capita income in thousands of dollars	50.21	14.39	26.79	179.74
	mean	sd	min	max
Panel B: 3 Digit Zip Code Level Data				
Workplace drug test marijuana positivity rate	2.01	0.94	0.00	6.50
Log of network exposure	26.59	0.46	25.63	28.41
Standardized min. distance to a legalized state	0.00	1.00	-1.29	3.91
Neighboring a legalized state	0.11	0.32	0.00	1.00
Having a direct flight to a legalized state	0.92	0.27	0.00	1.00
Unemployment rate	5.11	1.25	2.53	11.08
Per capita income in thousands of dollars	52.68	16.36	28.27	216.98

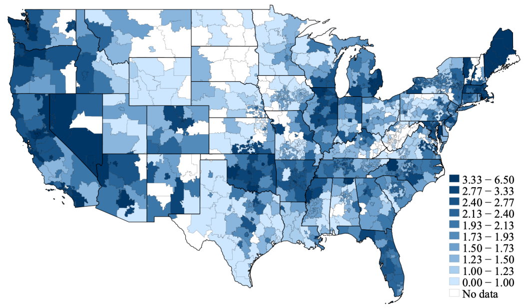
Notes: In Panel A, observation is a substate in NSDUH. There are 245 substates in our sample. In Panel B, observation is a 3-digit zip code. There are 753 3-digit zip codes in our sample. The table presents statistics from the latest available time period (2018 for panel A and 2020 for panel B; except the reported marijuana use, which is the three-year average from the 2016-2018 waves of the NSDUH Public Substate Report).



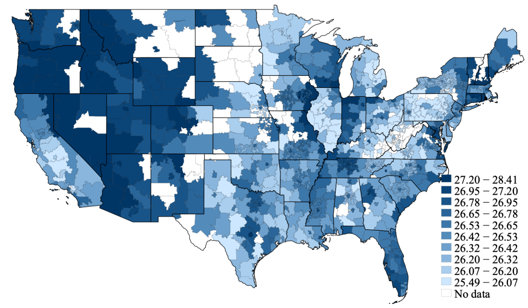
(a) Incidence of Marijuana Use in the Past Year at substate level, 2016-2018



(b) Network Exposure to states with legal use at substate level (in logs), 2018



(c) Workplace marijuana positivity rate at 3 digit zip code level, 2018-2020



(d) Network Exposure to states with legal use at 3 digit zip code level (in logs), 2020

Figure 1: Geographic Distribution of our Outcome Variables and Network Exposure Measure

Notes: Our main samples exclude substates and 3 digit zip codes in Alaska, and Hawaii. Additionally, our substate level analysis excludes two substates in Georgia and all substates from North Carolina because they changed the substate borders and are not comparable across NSDUH waves. The 3 digit zip code analysis includes localities in these states.

4 Empirical strategy

Fixed-effects regression. Using the network exposure measures we estimate the regression where the probability of marijuana use in locality i in state s in period t equals:

$$\begin{aligned} MarijuanaUse_{it} = \beta \log(Exposure_{it}) + X_{it}'\alpha + LocalityFE_i \\ + StateYearFE_{st} + \varepsilon_{it} \end{aligned} \tag{2}$$

where the outcome variable $MarijuanaUse_{it}$ is either the percentage of the population that reported using marijuana in the past year or the workplace drug test marijuana positivity rate. $log(Exposure_{it})$ is the main variable of interest and is defined by Equation (1). X_{it}' is a vector of time-varying covariates, such as, the unemployment rate and the logarithm of per capita income. The regression includes fixed effects for each locality, $LocalityFE_i$. It also includes state-year pair fixed effects, $StateYearFE_{st}$. The state-year fixed effects control for state-level changes, for example, when a state itself legalizes marijuana. As the state-year fixed effects are included, we are effectively comparing localities within the same state-year. We cluster standard errors at the state level.

Figure A.3 in the Online Appendix illustrates our identification strategy. Prior to 2012, when the first state legalized marijuana, the network exposure was uniformly zero across all substates. We display the network exposure at the substate level in Florida for the year 2018 in Panel a. This network exposure was imputed based on the population-weighted network exposure to states that legalized marijuana between 2012 and 2017. Panel b presents the last year's marijuana usage from NSDUH between 2016 and 2018. Our regression would show us whether there is a relationship between panel a and panel b. In one of the robustness checks, we also impute the network exposure (panel a) as a 3-year average, and the results remain robust.

Some might worry that the level of SCI reflects the distance between two localities, and we are capturing the behavior of two close localities which might experience similar shocks. For example, a substate that is neighboring Colorado, which legalized recreational marijuana in 2012, may experience an increase in marijuana use in 2012 because of its physical proximity and an increase in easy access to marijuana. The same substate could also have a high SCI with substates in Colorado because of its physical proximity. We might attribute the importance of geographic proximity to the SCI due to the omitted variable bias. Therefore, $ProximityControls_{ist}$ includes several proximity measures, such as the shortest distance to any state that legalized marijuana (standardized), an indicator for neighboring any legalized state, and accessibility via direct flights to legalized states.

As we include state-year fixed effects, we compare marijuana usage in two localities within the same state and year, where one locality experiences a larger increase in exposure to legalized marijuana via more connections to states that have legalized. Additionally, by including locality fixed effects, we account for potential differences in the level of consumption between localities.

Our identification specification relies on the assumption that conditional on various controls, in the absence of legalization, regions with varying degrees of connection to states with legalized marijuana would have shown similar trends in marijuana smoking behavior over time. We assess this assumption through several approaches. Firstly, we investigate which baseline characteristics are correlated with high exposure to marijuana legalization. Our analysis in the Online Appendix Table A.2 reveals that areas with greater exposure to legalization tend to have higher levels of education, a larger proportion of Democratic voters, and a higher share of the black population. These differences in characteristics may indicate distinct trajectories for these regions. To address this concern, our specifications include a broader set of baseline characteristics (including the share of Democrats, share with bachelor's degrees, percent of Latinos, percent of foreign-born residents, percent with limited English proficiency, and percent aged 65 and above, all measured in 2010) and allows for diverging time trends based on these characteristics. The empirical specification is as follows:

$$
\begin{aligned}
MarijuanaUse_{ist} = \beta \log(Exposure_{ist}) + \alpha X'_{ist} \\
+BaselineChar_i \cdot t + ProximityControls_{ist} \\
+ LocalityFE_i + StateYearFE_{st} + \varepsilon_{ist}
\end{aligned}
\tag{3}
$$

Secondly, while we are aware of all state-level marijuana legalization laws up to 2023, we only analyze outcomes up to the year 2018 (NSDUH) and 2020 (Quest). Hence, we can utilize future law passing and the exposure between last year of the data to 2023 as a control. This serves as an additional control for some unobserved differences between the high-exposure and low-exposure localities.

Thirdly, in the Online Appendix, we analyze whether pre-legalization trends in marijuana use are correlated with changes in exposure. Figure A.4a presents scatter plots of residualized (after partialling out state fixed effects) changes in marijuana use in the period between 2004-2012 (prior to the first marijuana legalization in 2012) against residualized (after partialling out state fixed effects) changes in exposure between 2012 and 2018 (during the period of incremental state-level marijuana legalization). Prior to the first marijuana legalization, this correlation is small and not statistically significant at the 10 percent level. In contrast, Figure A.4b shows a large, statistically significant positive correlation between changes in marijuana use and changes in network exposure between 2012 and 2018. Similar figures are presented in Figures A.4c and A.4d using workplace marijuana positivity rate. In the residualized graph partiallying out only state fixed effects without additional controls, Figure A.4c shows a statistical significant correlation, the slope is only one-third the size compared to the post-legalization period in Figure A.4d. Given that there is a very small positive correlation between changes in outcome in the pre-2012 period and network exposure, in conjunction with what we find in Table A.2 and Columns 3 and 4 of Table 2, we can conclude that network exposure to legalization may be associated with some underlying differences in characteristics, but these differences are not the main drivers of our results.

In our main analysis, we present two sets of results. First, we present results from all the time periods (the standard difference-in-difference specifications), then we estimate regression (2) using data from only three periods: two before the legalization and one after some states have legalized. Specifically, the three periods are 2002–2004, 2010–2012, and

2016-2018 for NSDUH and 2007–2009, 2010–2012, and 2018-2020 for Quest data. The benefit of using less granular data is that now we have only a single period with treatment. If instead, we were to use data from all time periods, then we would have a staggered roll-out of the treatment, where states legalize marijuana in different time periods. Recent econometrics literature has shown that standard difference-in-differences regressions often do not provide valid estimates of the average treatment effect when treatment timing is staggered (for an overview of the literature, see a recent survey by de Chaisemartin and D'Haultfœuille (2023)). Our approach avoids analyzing staggered treatment and is likely to give more robust estimates. In the Online Appendix, we provide heterogeneity-robust estimates developed by de Chaisemartin and D'Haultfœuille (2020). We discuss this in more detail in the next section.

Event study. To analyze changes over time and evaluate whether high-exposure localities had similar trends before legalization as the low-exposure localities, we estimate an event study. To estimate an event study, we consider each state's legalization as a separate event and estimate its impact on all localities in other states. The following specification measures the effect of network exposure from the legalization in state k on localities i in state s at time t:

$$\begin{aligned} MarijuanaUse_{it} = \sum_{\tau=-1}^{1} \beta_\tau \log\left(SCI_{ik} \times 1[\tau\ PeriodFromLegalizationInState\ k]\right) \\ + LocalityFE_i + X'_{st} + StateYearFE_{st} + \varepsilon_{it} \end{aligned} \tag{4}$$

where SCI_{ik} is the Social Connectedness Index measuring the relative probability in localities i having friends in state k. The Facebook dataset does not include substate-to-state Social Connectedness Index. Therefore, we calculate the locality i to state k Social Connectedness Index from the county-to-county or zipcode-to-zipcode indexes using the population-weighted average. More details of the imputation are in the Online Appendix Section C. The results from running this equation for each implementing state are in Figures A.5 and A.6.

In our main specification, instead of estimating regression (4) separately for each state that legalized, we stack the data and estimate the average effect across legalization events. Event study estimates separately for each state that legalized are presented in the Online Appendix. The stacked event study framework has been used, for example, by Autor, Donohue and Schwab (2006); Cengiz et al. (2019); Deshpande and Li (2019) and in a more similar setting by Wilson (2022). An observation then is locality i in state s time period t, and state legalization event k. We estimate the following event study regression:

$$\begin{aligned} MarijuanaUse_{istk} = \sum_{\tau=-1}^{1} \beta_\tau \log\left(SCI_{ik} \times 1[\tau \ PeriodFromLegalizationInState \ k]\right) \\ + LocalityEventFE_i + StateEventYearFE_{stk} + \varepsilon_{itk} \end{aligned} \tag{5}$$

where we include locality fixed effects $LocalityEventFE_i$, and state-year-event triplet fixed effects. To have a balanced panel, it is not possible to estimate the regression across legalization events that are either at the beginning or end of the time period. Therefore, our analysis focuses on analyzing the impact of those states that implemented the laws between 2012 and 2016. We conduct the analysis with 3 periods: 2002-2004, 2010-2012, and 2016-2018 (for NSDUH) and 2007-2009, 2010-2012, 2018-2020 (for Quest). As before, the benefit of using less granular data is that now we do not have a staggered roll out of treatment, which has its issue as suggested by Sun and Abraham (2021). The simple proposed solution of comparing the treatment group to an unaffected group would not work in our study since there is no unaffected group. All localities are potentially affected by every state implementation as long as there is some connection.

Threat to identification. A potential concern arises regarding the construction of the network exposure measure using the 2021 Social Connectedness Index (SCI). Most recreational cannabis legalization occurred between 2015 and 2020, raising the possibility of legalization influencing the network. However, it is important to note that Facebook claims the SCI to be highly stable over time. The correlation between the county-to-county SCI for 2016 and 2021 in our sample is 0.9. We note that given its stability, we are taking the 2021 SCI as a baseline SCI. The variation in the main variable of interest, network exposure over time, is mainly driven by the timing of marijuana legalization and the connections to the state that has enacted the law. Furthermore, in the robustness check section, we find that the results for marijuana use remain positive and significant at 5% when utilizing either the 2016 or 2021 county-to-county SCI to construct the exposure measure. This further confirms the stability of the SCI over time and strengthens our confidence in the results.

5 Main results

Evidence of the direct effect of marijuana legalization on the state that legalized. Before delving into our main analysis of the network's role, it is crucial to confirm the direct impact of marijuana legalization on marijuana use within the states that have legalized it. While Hollingsworth, Wing and Bradford (2022) demonstrated this relationship, it is necessary to validate it ourselves since we utilize aggregated marijuana data at the substate level, unlike their state-level analysis.

To accomplish this, we employ a panel data fixed effects regression model, where the outcome variable is the percentage of the population aged 18 and above who reported marijuana use in the past year. The key variable of interest indicates the legalization status of marijuana use within a state. Each regression incorporates substate fixed effects and year-Census region fixed effects or year fixed effects to account for the variability across different substates and over time.

In column 1 of Table A.3, we observe that the legalization of marijuana use does indeed lead to a significant increase of 3.6 percentage points in marijuana use within the states that have legalized it. This represents a substantial change, considering that prior to legalization, only 15 percent of individuals aged 18 and above reported using marijuana in the past year. Hence, the legalization resulted in an approximate 24% increase in yearly usage. In Panel B, we present the results using Quest dataset. Legalization of marijuana use increases the marijuana positivity rate by 0.66 percentage points, which is about a 32% increase.

Panel data fixed effects results. In Table 2, we add the additional controls one by one across columns. Columns 1 and 5 present the estimated results from Equation (2). In columns 2 and 3, we present results from Equation (3). A one standard deviation increase in exposure to states with legalized use corresponds to a roughly one-quarter of a standard deviation increase in marijuana self-reported use and one-sixth of a standard deviation increase in workplace marijuana positivity rates.

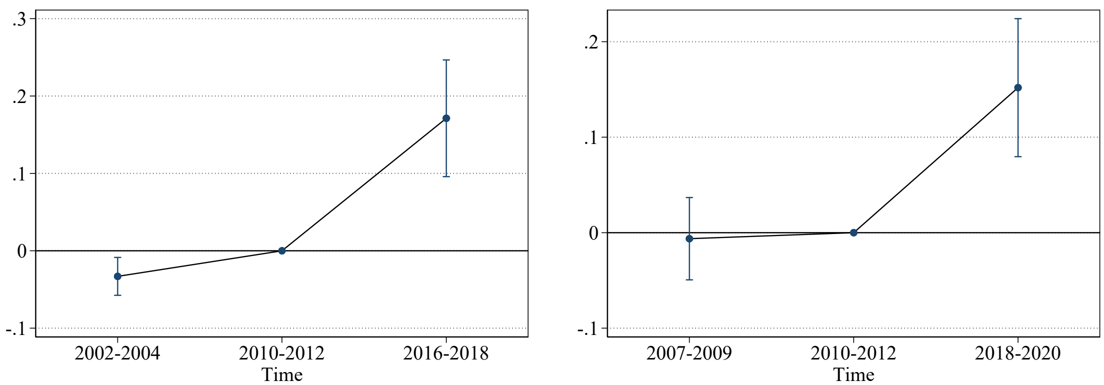
Event study results. To analyze changes over time and assess whether substates with varying levels of exposure to legalized states had similar trends before legalization, we conduct an event study. Figure 2 presents the event study estimates. As discussed in Section 4, we present the event study results in two ways. Panels a and c utilize all legalized states in the analysis and we only examine the 3 periods (two periods prior to any legalization, and one period post legalization). We only examine the events in which states legalized marijuana use between 2012 and 2016. In these panels, the trends in marijuana consumption prior to out-of-state marijuana legalization were comparable in areas with both high and low social network exposure. This finding suggests that the larger impact observed in high-exposure areas is not due to divergent trends before legalization. Moreover, Figures A.5 and A.6 in the Online Appendix demonstrates that the event study estimates are similar when each event (state passing marijuana legalization law) is examined separately.

Table 2: Indirect impact of exposure to states that have legalized marijuana on marijuana outcomes

	Panel A: Marijuana Use in Past Year							
	All Years: 2004-2018				3 Periods: 2004, 2012, 2020			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log. exposure to	0.955***	1.043***	1.290***	1.430***	1.276***	1.338***	1.727***	2.159***
states with legal use	(0.258)	(0.297)	(0.349)	(0.372)	(0.423)	(0.470)	(0.605)	(0.793)
Future Exposure				-0.302				-0.714
				(0.351)				(0.498)
Year-state FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Substate FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Proximity Measures	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Baseline X Trend	No	No	Yes	Yes	No	No	Yes	Yes
Num of Substates	245	245	245	245	245	245	245	245
Observations	1960	1960	1960	1960	735	735	735	735
	Panel B: Workplace Positivity Rate for Marijuana							
	All Years: 2007-2020				3 Periods: 2009, 2012, 2020			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log exposure to	0.241***	0.314***	0.176**	0.130*	0.431***	0.504***	0.395***	0.347**
states with legal use	(0.071)	(0.074)	(0.071)	(0.077)	(0.098)	(0.117)	(0.133)	(0.132)
Future Exposure				0.093				0.079
				(0.074)				(0.097)
Year-State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3 digit zip code FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Proximity Measures	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Baseline X Trend	No	No	Yes	Yes	No	No	Yes	Yes
Num of 3 digit zip code	753	753	753	753	753	753	753	753
Observations	10542	10542	10542	10542	2259	2259	2259	2259

Notes: All columns are the impact of exposure to other states that have legalized marijuana. Columns 1 and 5 reflect estimates of *beta* from Equation (2). All regressions include the unemployment rate and the logarithm of per capita income. Proximity measures include standardized minimum distance to any legalized state, an indicator for whether there was a direct flight between the locality to any legalized state, and an indicator variable for whether the locality is neighboring any legalized state. Baseline characteristics-specific time trends are described in Equation (3). Columns 4 and 8 include an additional future exposure measure. It captures the eventual network exposure (from either 2018 to 2023 (in Col 4) or 2020 to 2023 (in Col 8). Columns 1-4 use data from all periods while columns 5-8 only use data from only 3 periods: two periods prior to any marijuana legalization and one period post marijuana legalization.

Robustness and permutation test. Below, we summarize the analyses conducted to assess the sensitivity of our results to alternative functional forms, controls, clustering, and samples. These findings are presented in Figure 3. Each bar represents a point estimate and the 95% confidence interval derived from a separate regression, analogous to Equation (3). The coefficient from column 7 of Table 2 marks the red line in this figure. In Panel A of the NSDUH dataset analysis, we incorporate bootstrapped standard errors (Estimation 1). In



(a) Marijuana use in past year

(b) Workplace drug test

Figure 2: Event study: the impact of the probability of having out-of-state friends exposed to a new state marijuana legalization on marijuana use in the past year (Figure 2a), and on positivity rate for marijuana in workplace drug tests (Figure 2b)

Notes: Figure 2a uses data of three time periods: 2002-2004, 2010-2012, and 2016-2018; similarly, Figure 2b uses data of 2007-2009, 2010-2012, and 2018-2020. The data is stacked, each area N times where N is the number of states that legalized from December 2012 until December 2015. (Results for each state legalization separately are presented on Figures A.5 and A.6.) There are state-event-year fixed effects and locality-event fixed effects, unemployment rates and income per capita in all specifications. Standard errors are clustered at either substate level for marijuana use or 3-digit-zip-code level for workplace drug test results.

Estimation 2, we deliberately omit all states that have previously legalized marijuana. In Estimation 3, we exclude those states that have the below median distance to a legalized state. The median distances are approximately 290 miles in 2018 (Panel a) and 190 miles in 2020 (Panel b). These coefficients from the subsample are not statistically different from the full sample specification (highlighted in the red line), suggesting that exposure from distant peers has a similar impact to that from close peers. This is an important finding, as it indicates that in the modern age, distance between people matters less; distant peers can still exert significant influence, likely through smartphones, even if they are more than 200 miles away.

We also explore alternative specifications for the key variable of interest: using a 3-year average of exposure (Estimation 4), namely raw variable of *Network exposure* (Estimation 5), as well as standardizing *Network exposure* (Estimation 6). In our main analysis, we use the legalization of marijuana use as the primary law change to estimate exposure. Here, we provide estimates using the legalization of marijuana licensed sales (Estimation 7) to determine exposure.

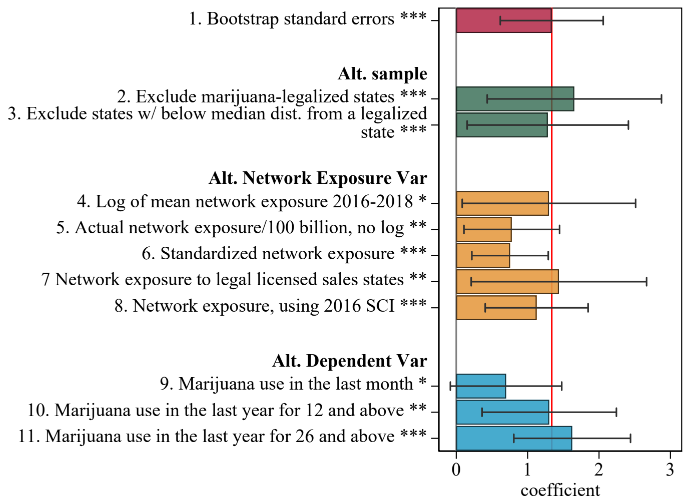
In our main specification, the social connectedness index used was constructed based on Facebook connections in 2021. However, considering the possibility of friendship formation as a result of marijuana legalization and potential reverse causality, we obtained the 2016 version's county-to-county SCI. Consequently, we constructed substate-level SCI in 2016 and used it with the NSDUH data. The results from using the 2016 SCI are reported in Estimation 8.[8] Furthermore, in the NSDUH dataset, each respondent reports whether they used marijuana in the previous month or not. We use last month's marijuana use as a dependent variable (Estimation 9). Finally, we conduct the analysis by self-reported marijuana use above age 12 and above and 26 and above (Estimations 10 and 11). Regarding the workplace drug test data, for each 3-digit zip code in each year, the dataset reports a range of the positivity rate for three-quarters of the observations. However, for one-quarter of the time, the raw data only provides a single number for the positivity rate. In our main specification, we consider this single number as the lower bound. In Figure 3 Panel b, Estimations 8 and 9, we display the coefficients obtained when we treat the single number as the median or the upper bound, and use it as the dependent variable, respectively.

Recent literature has highlighted that the standard two-way fixed effects estimator may be biased when treatment effects exhibit heterogeneity and treatment is staggered (for an overview, refer to a recent survey by de Chaisemartin and D'Haultfœuille (2023)). First, to assess the robustness of the estimates in Table 2, in the Online Appendix (Table A.4), we also present the heterogeneity-robust estimates proposed by de Chaisemartin and D'Haultfœuille (2020) using the data where treatment is a continuous variable. The point estimates remain similar to our main results in Table 2. Second, in the Online Appendix (Table A.5), we also present the estimates from a regression where the treatment variable is binary (below or above median exposure) and treatment is not staggered (only a single period after treatment). When the treatment is not staggered and is binary, the estimates are not subject to the concerns raised above because then the standard difference-in-differences estimator is equivalent to the heterogeneity robust estimator proposed by de Chaisemartin and D'Haultfœuille (2023).

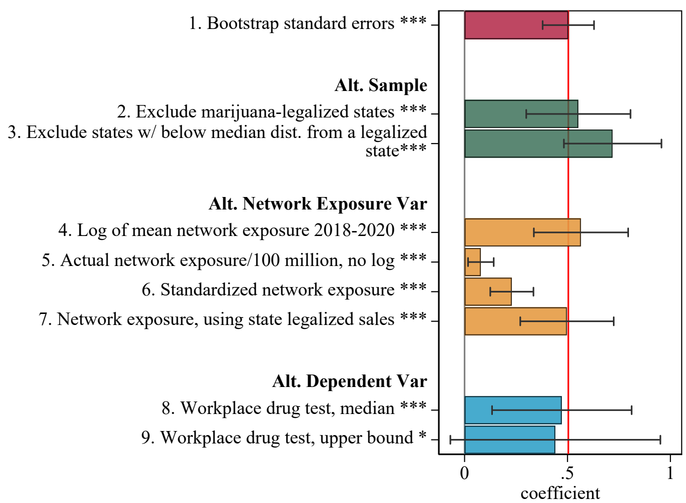
Last, we perform permutation tests. Since our key variable of interest is network exposure to marijuana legalization, we shuffle $LogNetworkExposure$ within a given state and year 500 times and run the regression specification the same as Equation (2).[9] We present the distribution of coefficient estimates and t-tests results in Figure A.7. The red bar in each panel indicates the corresponding statistics from column 6 from Table 2. One can see that our estimates are far from the estimates generated from randomly assigned network exposure.

[8]Since we do not have the 2016 zip code to zip code SCI, we cannot generate analogous SCI for Quest data. Additionally, Facebook changed the SCI scale between 2016 and 2021, making the coefficients on 2016 and 2021 SCI not directly comparable. Therefore, we should not compare the magnitude of the coefficients between 2016 and 2021 SCI.

[9]In this exercise, we exclude those states that have fewer than 4 substates. Otherwise, often it would shuffle back to the actual assignment.



(a) Marijuana Use



(b) Workplace Drug Test

Figure 3: Robustness checks

Notes: Each bar presents a coefficient and its 95% confidence interval from a separate regression. All regressions are based on the specification in Equation (3), including state-year FE, locality FE, geographic proximity controls, unemployment, local income per capita, and baseline characteristics-specific time trends. Est. 1 reports standard errors from bootstrapping 500 times. Est. 2 uses an alternative sample with only localities where marijuana is still illegal. Est. 3 excludes those states that have below median distance from a legalized state. The cutoffs are 290 miles in Panel A in 2018 and 190 miles in Panel B in 2020. Est. 4 uses the natural log of 3 years average of network exposure (e.g. 2016–2018) instead of just the last year (e.g. 2018). Est. 5 uses the actual network exposure. Est. 6 uses standardized network exposure. Est. 7 uses marijuana licensed sales legalization to impute the network exposure. Est. 8 in Panel (a) uses 2016 SCI to impute network exposure. Est. 9, 10, and 11 in Panel (a) use alternative dependent variables: last month's use, marijuana use for age 12 and above, and marijuana use for age 26 and above, respectively. Est. 8 and 9 in Panel (b) use the median and upper bound of the workplace drug test data as alternative dependent variables. The red bar indicates the coefficient estimates from column 6 in Table 2.

It confirms the importance of the network exposure variable we use in our analysis.

Heterogeneity by age group. In the results not presented here, we also examined marijuana use across different age groups, as reported by NSDUH. Specifically, NSDUH provides data on marijuana use among individuals aged 12 to 17, aged 18 to 25, separately. While the findings for these subgroups still indicate positive associations, their estimates are less precise, with one subgroup yielding a p-value of 0.69 and the other 0.17.

Several potential reasons may account for these coefficients not achieving the same level of statistical significance as the main specifications for ages 18 and above. Firstly, we encountered a limitation in terms of sample size for these subgroups. Whenever the number of observations from a specific substate is insufficient, NSDUH does not report the corresponding mean. Therefore, we lost nearly one-third of our sample in the age group 18 to 25. Second, the literature on the direct impact of recreational marijuana legalization on adolescents in legalized states presents mixed findings. Anderson et al. (2019); Coley et al. (2021) suggest that the effect of recreational marijuana use on increasing adolescent marijuana use in these states is inconclusive. Another possibility pertains to the fact that Facebook SCI may not be a good proxy for the younger age group's social network. According to Pew Research Center (2022), the top 3 social media platforms for teens between age 13 and 17 are Youtube, Tiktok, and Instagram. Consequently, the network exposure employed in this study may not be as relevant for this particular age bracket. To illustrate this, consider an extreme scenario where none of the individuals aged 12 to 18 have access to smartphones. In such a case, we would not expect these social networks to be as influential as they might be among smartphone users.

Mechanism. To learn about the mechanism of how exposure to legalization via social network affects marijuana consumption, we isolate the variation in the social network that is due to geographic characteristics. Our goal is to understand the remaining importance of social network when we flexibly control for geographic characteristics. The exercise is motivated by the literature that has shown that geographic variables are strongly correlated with social network (Bailey et al., 2018*b*).

Our approach uses the following steps and is described in more detail in Appendix B. In the first step, we flexibly model the social connectedness index as a function of geographic characteristics (physical distance). The estimates from the first step are used to calculate the predicted (based on geographic characteristics) exposure according to Equation (1). We then use the control function approach and the predicted exposure is used as an instrument for the actual exposure variable.

When controlling for geographic characteristics in a flexible way, then in the NSDUH data, there is not much relevant variation left in the social network measure (Table B.3). Our main regressions in Table 2 also controlled for geographic proximity to legalized states and found that the exposure via social network still had a significant effect. However, it was a more restrictive specification, in terms of being dependent on specific functional forms, compared to the more flexible specification in Appendix B.

We conclude that the social network measure is a nuanced measure of connectedness that is partly explained by geographic variables. Therefore, attributing exposure effects solely to online social networks would be misleading. While we cannot rule out that the impact of exposure to legalization via Facebook connectedness is driven by some form of geographic connections, Facebook connectedness provides another approach to model spatial externalities.

Discussion. Social networks' impact is notoriously difficult to identify. Social connections are likely to form with areas that are more geographically connected either via road or flight networks. When a geographically connected state legalizes marijuana, does marijuana use increase because it is easy to go to buy from there or because of social connections? In this paper, we think of the impact of social networks in a broader sense. In the main regressions, we control for the geographic proximity to states that have legalized and interpret the remaining effect as the impact of social connections.

In our view, this SCI measure serves as a representation of both online and offline interconnectedness among any two localities within the United States. This perspective is reinforced by the findings in the study by Bailey et al. (2018*b*), which suggest that individuals typically add Facebook friends with whom they have real-world connections. Consequently, we cannot definitively attribute our results solely to the effects stemming from "likes" and comments on Facebook. However, the persistence of connections from long distances (greater than 190 miles) in influencing our behaviors underscores the role of social media and smartphones in our decision-making processes. These platforms serve as our primary means of connecting with friends and acquaintances who are geographically distant, highlighting their significance in shaping our choices.

Our results imply that studies estimating the direct impact of policies using a standard difference-in-differences approach without taking into account the externalities underestimate the direct effect of the policy. For example, without taking into account the externalities, the effect of marijuana legalization in the legalizing state is 3.9 percentage points (according to column 3 of panel A of Table A.3). Indeed, if there were no externalities to other states, then a 3.9 percentage points increase would be the total direct effect. But the externality to the above median connected areas is about 0.4 percentage points (column 5 Table A.5). Suppose for simplicity that in less connected areas, there were no externalities. Then back-of-the-envelope calculations suggest that the direct effect of the policy is instead about 4.1 percentage points. Hence, without taking into account the externality, we underestimated the direct effect by about five percent.

6 Conclusion

In conclusion, this paper investigates the spatial externalities of marijuana legalization on marijuana use. Our findings indicate that connections to states where recreational marijuana use is legalized significantly impact marijuana use and workplace drug testing positivity rates in areas where marijuana use remains illegal.

We utilize the Facebook Social Connectedness Index, which measures the strength of connections between different geographic areas based on Facebook friendship ties, as a measure of exposure to legalized marijuana among Facebook friends. Specifically, within a given state, a one standard deviation increase in exposure to states with legalized use corresponds to roughly a one-quarter standard deviation increase in reported marijuana use and a one-sixth standard deviation increase in workplace marijuana positivity rates.

Overall, our findings highlight the importance of considering spatial externalities in policy. These connections can lead to increased adoption of risky behaviors, such as marijuana use, even in areas where marijuana remains illegal. Moreover, our subsample analysis reveals that influence from distant peers is as strong as that from nearby peers. This significant finding suggests that in the modern era, physical distance between individuals is less important. Even when separated by long distances, people can still have a powerful impact on each other.

Disclaimer During the preparation of this work the authors used ChatGPT to improve the language and to check grammar and spelling. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Abouk, Rahi, Mohammad Mansouri, and David Powell. 2024. "Can Access to Non-Opioid Pain Management Therapies Reduce Overdose Rates? Evidence from Medical Cannabidiol Laws." SSRN 4683257.

Adda, Jérôme, Brendon McConnell, and Imran Rasul. 2014. "Crime and the Depenalization of Cannabis Possession: Evidence from a Policing Experiment." *Journal of Political Economy*, 122(5): 1130–1202.

Anderson, D. Mark, Benjamin Hansen, and Daniel I. Rees. 2013. "Medical Marijuana Laws, Traffic Fatalities, and Alcohol Consumption." *The Journal of Law and Economics*, 56(2): 333–369.

Anderson, D Mark, Benjamin Hansen, Daniel I Rees, and Joseph J Sabia. 2019. "Association of marijuana laws with teen marijuana use: new estimates from the youth risk behavior surveys." *JAMA pediatrics*, 173(9): 879–881.

Autor, David H, John J Donohue, III, and Stewart J Schwab. 2006. "The Costs of Wrongful-Discharge Laws." *The Review of Economics and Statistics*, 88(2): 211–231.

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel. 2021. "International trade and social connectedness." *Journal of International Economics*, 129: 103418.

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel. 2019. "House Price Beliefs And Mortgage Leverage Choice." *The Review of Economic Studies*, 86(6): 2403–2452.

Bailey, Michael, Patrick Farrell, Theresa Kuchler, and Johannes Stroebel. 2020. "Social connectedness in urban areas." *Journal of Urban Economics*, 118: 103264.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018*a*. "Social Connectedness: Measurement, Determinants, and Effects." *Journal of Economic Perspectives*, 32(3): 259–280.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018*b*. “Social Connectedness: Measurement, Determinants, and Effects.” *Journal of Economic Perspectives*, 32(3): 259–280. Facebook Data for Good Program, Social Connectedness Index (SCI). https://dataforgood.facebook.com/, accessed June 7, 2022.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018*c*. “The Economic Effects of Social Networks: Evidence from the Housing Market.” *Journal of Political Economy*, 126(6): 2224–2276.

Bradford, Ashley C., and W. David Bradford. 2018. “The Impact of Medical Cannabis Legalization on Prescription Medication Use and Costs under Medicare Part D.” *The Journal of Law and Economics*, 61(3): 461–487.

Card, David, and Laura Giuliano. 2013. “Peer Effects and Multiple Equilibria in the Risky Behavior of Friends.” *The Review of Economics and Statistics*, 95(4): 1130–1149.

Carrell, Scott E., Richard L. Fullerton, and James E. West. 2009. “Does Your Cohort Matter? Measuring Peer Effects in College Achievement.” *Journal of Labor Economics*, 27(3): 439–464.

Carrieri, Vincenzo, Leonardo Madio, and Francesco Principe. 2020. “Do-It-Yourself medicine? The impact of light cannabis liberalization on prescription drugs.” *Journal of Health Economics*, 74: 102371.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. “The Effect of Minimum Wages on Low-Wage Jobs.” *The Quarterly Journal of Economics*, 134(3): 1405–1454.

Choi, Anna, Dhaval Dave, and Joseph J. Sabia. 2019. “Smoke Gets in Your Eyes: Medical Marijuana Laws and Tobacco Cigarette Use.” *American Journal of Health Economics*, 5(3): 303–333.

Clark, Andrew E., and Youenn Lohéac. 2007. ““It wasn’t me, it was them!” Social influence in risky behavior by adolescents.” *Journal of Health Economics*, 26(4): 763–784.

Coley, Rebekah Levine, Claudia Kruzik, Marco Ghiani, Naoka Carey, Summer Sherburne Hawkins, and Christopher F Baum. 2021. “Recreational marijuana legalization and adolescent use of marijuana, tobacco, and alcohol.” *Journal of Adolescent Health*, 69(1): 41–49.

Cutler, David M., and J. Travis Donahoe. 2024. "Thick Market Externalities and the Persistence of the Opioid Epidemic." NBER 32055.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–2996.

DeCicca, Philip, Donald Kenkel, and Feng Liu. 2013. "Excise tax avoidance: The case of state cigarette taxes." *Journal of Health Economics*, 32(6): 1130–1141.

Deshpande, Manasi, and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs." *American Economic Journal: Economic Policy*, 11(4): 213–248.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2023. "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey." *The Econometrics Journal*, 26(3): C1–C30.

Dong, Xiuming. 2022. "Recreational Marijuana Sales Legalization and Monday Work Injury Claims." *The B.E. Journal of Economic Analysis & Policy*, 22(1): 99–121.

Dragone, Davide, Giovanni Prarolo, Paolo Vanin, and Giulio Zanella. 2019. "Crime and the legalization of recreational marijuana." *Journal of Economic Behavior & Organization*, 159: 488–501.

Eisenberg, Daniel, Ezra Golberstein, and Janis L. Whitlock. 2014. "Peer effects on risky behaviors: New evidence from college roommate assignments." *Journal of Health Economics*, 33: 126–138.

Gavrilova, Evelina, Takuma Kamada, and Floris Zoutman. 2019. "Is Legal Pot Crippling Mexican Drug Trafficking Organisations? The Effect of Medical Marijuana Laws on US Crime." *The Economic Journal*, 129(617): 375–407.

Hansen, Benjamin, Keaton Miller, and Caroline Weber. 2020. "Federalism, partial prohibition, and cross-border sales: Evidence from recreational marijuana." *Journal of Public Economics*, 187: 104159.

Harding, Matthew, Ephraim Leibtag, and Michael F. Lovenheim. 2012. "The Heterogeneous Geographic and Socioeconomic Incidence of Cigarette Taxes: Evidence from Nielsen Homescan Data." *American Economic Journal: Economic Policy*, 4(4): 169–98.

Hollenbeck, Brett, and Kosuke Uetake. 2021. "Taxation and market power in the legal marijuana industry." *The RAND Journal of Economics*, 52(3): 559–595.

Hollingsworth, Alex, Coady Wing, and Ashley C. Bradford. 2022. "Comparative Effects of Recreational and Medical Marijuana Laws on Drug Use among Adults and Adolescents." *The Journal of Law and Economics*, 65(3): 515–554.

Jacks, David S, Krishna Pendakur, and Hitoshi Shigeoka. 2021. "Infant Mortality and the Repeal of Federal Prohibition." *The Economic Journal*, 131(639): 2955–2983.

Jacobi, Liana, and Michelle Sovinsky. 2016. "Marijuana on Main Street? Estimating Demand in Markets with Limited Access." *American Economic Review*, 106(8): 2009–2045.

Johansson, Per, Tuomas Pekkarinen, and Jouko Verho. 2014. "Cross-border health and productivity effects of alcohol policies." *Journal of Health Economics*, 36: 125–136.

Kelly, Elaine, and Imran Rasul. 2014. "Policing cannabis and drug related hospital admissions: Evidence from administrative records." *Journal of Public Economics*, 112: 89–114.

Kremer, Michael, and Dan Levy. 2008. "Peer Effects and Alcohol Use among College Students." *Journal of Economic Perspectives*, 22(3): 189–206.

Lovenheim, Michael F., and Joel Slemrod. 2010. "The fatal toll of driving to drink: The effect of minimum legal drinking age evasion on traffic fatalities." *Journal of Health Economics*, 29(1): 62–77.

Lundborg, Petter. 2006. "Having the wrong friends? Peer effects in adolescent substance use." *Journal of Health Economics*, 25(2): 214–233.

Marie, Olivier, and Ulf Zölitz. 2017. ""High" Achievers? Cannabis Access and Academic Performance." *The Review of Economic Studies*, 84(3): 1210–1237.

McMichael, Benjamin J., R. Lawrence Van Horn, and W. Kip Viscusi. 2020. "The impact of cannabis access laws on opioid prescribing." *Journal of Health Economics*, 69: 102273.

Merriman, David. 2010. "The Micro-geography of Tax Avoidance: Evidence from Littered Cigarette Packs in Chicago." *American Economic Journal: Economic Policy*, 2(2): 61–84.

Mäckle, Kai, and Stefan Ruenzi. 2023. "The Social Transmission of Non-Infectious Diseases: Evidence from the Opioid Epidemic." SSRN 4104796.

Powell, David, Rosalie Liccardo Pacula, and Mireille Jacobson. 2018. "Do medical marijuana laws reduce addictions and deaths related to pain killers?" *Journal of Health Economics*, 58: 29–42.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *The Quarterly Journal of Economics*, 116(2): 681–704.

Sacerdote, Bruce. 2011. "Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In *Handbook of the Economics of Education*. Vol. 3, , ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 249–277. Elsevier.

Sovinsky, Michelle, Liana Jacobi, Alessandra Allocca, and Tao Sun. 2023. "More than Joints: Multi-Substance Use, Choice Limitations, and Policy Implications." CEPR Discussion Paper 18693.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of econometrics*, 225(2): 175–199.

Vannucci, Anna, Emily G. Simpson, Sonja Gagnon, and Christine McCauley Ohannessian. 2020. "Social media use and risky behaviors in adolescents: A meta-analysis." *Journal of Adolescence*, 79: 258–274.

Wen, Hefei, Jason M. Hockenberry, and Janet R. Cummings. 2015. "The effect of medical marijuana laws on adolescent and adult use of marijuana, alcohol, and other substances." *Journal of Health Economics*, 42: 64–80.

Williams, Jenny, and Anne Line Bretteville-Jensen. 2014. "Does liberalizing cannabis laws increase cannabis use?" *Journal of Health Economics*, 36: 20–32.

Wilson, Riley. 2022. "The Impact of Social Networks on EITC Claiming Behavior." *The Review of Economics and Statistics*, 104(5): 929–945.

A Online appendix: Additional figures and tables

Table A.1: Changes in recreational marijuana laws until the end of 2020

State	Legalized Recreational Cannabis	Licensed Sales
Washington	December 6, 2012	July 8, 2014
Colorado	December 10, 2012	January 1, 2014
Alaska	February 24, 2015	October 29, 2016
Washington, D.C.	February 26, 2015	
Oregon	July 1, 2015	October 1, 2015
California	November 9, 2016	January 1, 2018
Massachusetts	December 15, 2016	November 20, 2018
Nevada	January 1, 2017	July 1, 2017
Maine	January 30, 2017	October 9, 2020
Vermont	July 1, 2018	
Michigan	December 6, 2018	December 1, 2019
Illinois	January 1, 2020	January 1, 2020
Arizona	November 30, 2020	

Source: `https://en.wikipedia.org/wiki/Legality_of_cannabis_by_U.S._jurisdiction`, accessed July 24, 2024.

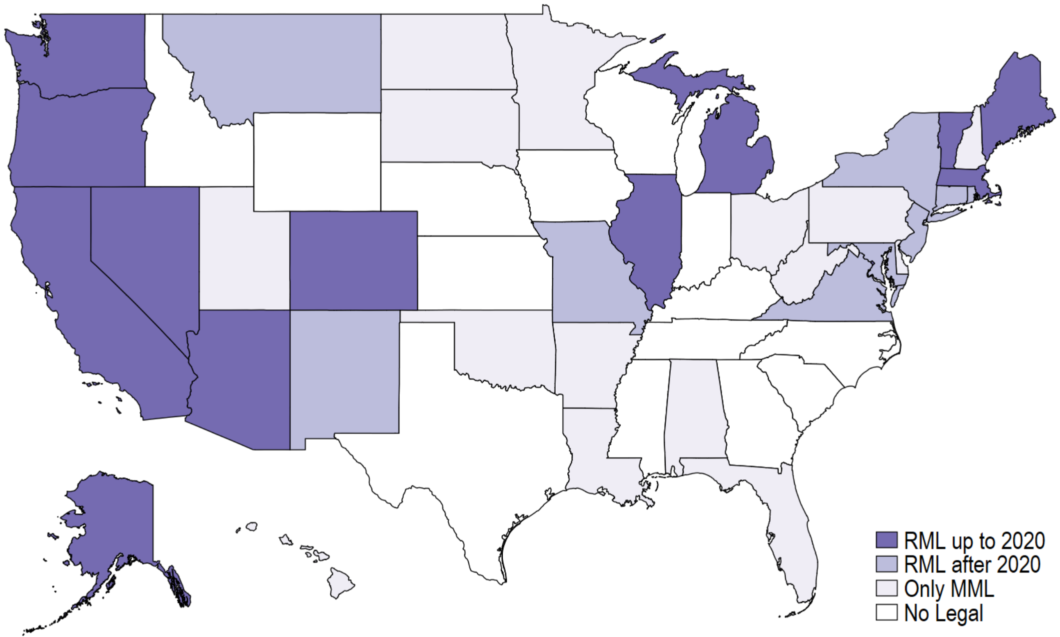


Figure A.1: Medical and Recreational Marijuana Laws

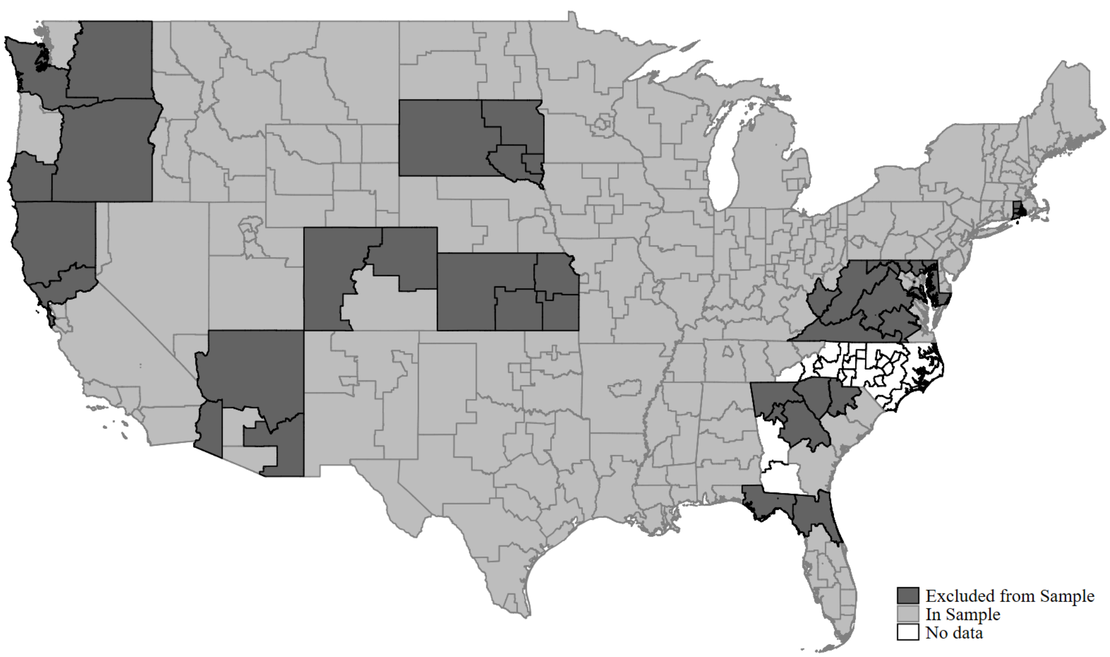
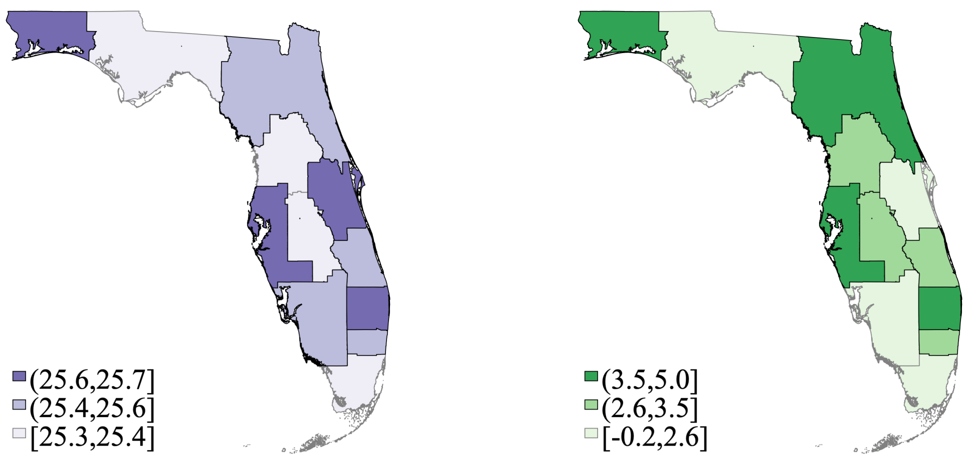


Figure A.2: Distribution of Substates

Notes: All substates in District of Columbia are excluded because there is no variation in the social conectedeness index at the substate level. D.C. only has one county. Two substates in Georgia were excluded. All substates from North Carolina were excluded from the sample.

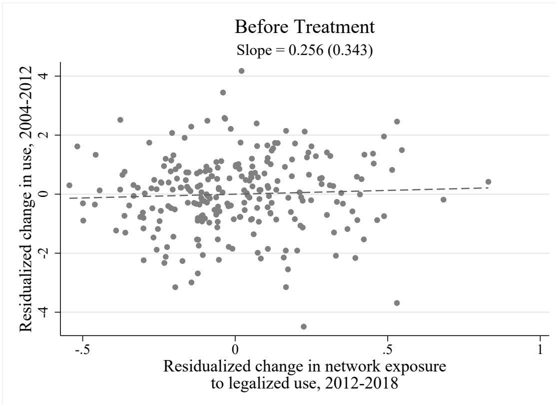


(a) Change in Network Exposure to Legalized States

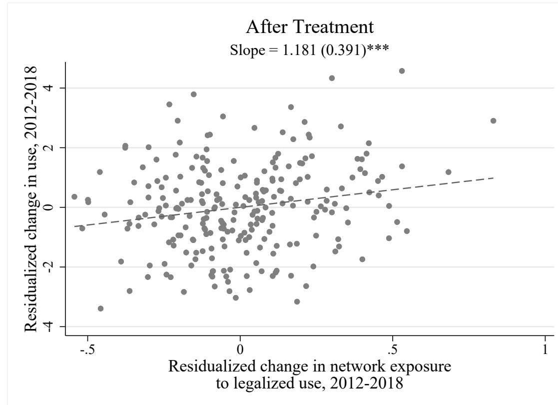
(b) Change in marijuana use during treatment

Figure A.3: **Florida**: Change in the network exposure to legalized use (in 2012–2018) vs change in marijuana use during treatment

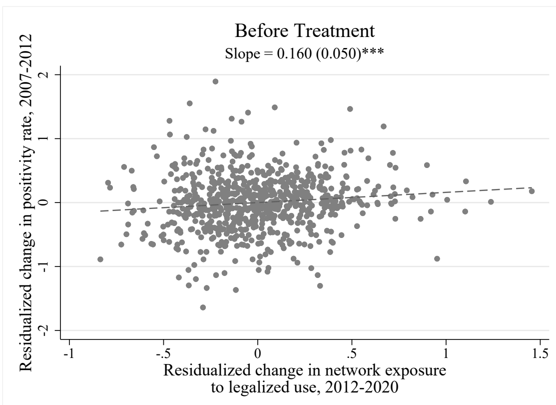
Each block represents a substate in NSDUH. Network exposure was zero in all substates before 2012. Panel a indicates variation in the natural log of network exposure at the substate level, and Panel b shows the marijuana use which is available from 2016-2018 wave of NSDUH public data.



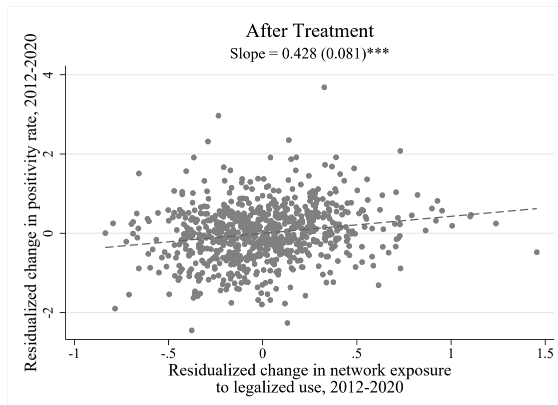
(a) Change in marijuana use before treatment



(b) Change in marijuana use during treatment



(c) Change in employee marijuana positivity rate before treatment



(d) Change in employee marijuana positivity rate during treatment

Figure A.4: Relationship between the exposure to legalization (in 2012–2020) and the change in yearly use before (a and c) and during treatment (b and d)

Notes: Each figure presents a scatter plot of the residualized change in yearly marijuana use or marijuana positivity rates (y-axes) on the residualized changes of the network exposure (x-axes). For each variable, we first take the residuals after controlling for state fixed effects. X-axes in all four figures measure the residualized change in network exposure to legalized use between 2012 and 2018. In Figure A.4a, the y-axis is the change in marijuana use between 2002-2004 and 2010-2012. In Figure A.4b, the y-axis is the residualized change in marijuana use between 2010-2012 and 2016-2018. In Figure A.4c, the y-axis is the residualized change in workplace marijuana positivity rates between 2007-2009 and 2010-2012. In Figure A.4d, the y-axis is the residualized change in workplace positivity rate between 2010-2012 and 2018-2020. The line of best fit is obtained from OLS regression. Its slope coefficient and standard error (in parentheses) are reported on the graph. Figure 1a and 1b utilize NSDUH-substate sample while Figure 1c and 1d utilize Quest data.

Table A.2: Network Exposure and Baseline Characteristics

	Log of Network Exposure		
Unemployment Rate	-0.092***	-0.051	-0.026
	(0.030)	(0.039)	(0.032)
Per Capita Income	0.007***	-0.000	-0.000
	(0.002)	(0.001)	(0.001)
Share with Bachelor's Degree		3.619***	3.916***
		(0.641)	(0.528)
Share of Presidential Democratic Vote		0.460*	0.402*
		(0.258)	(0.214)
Share Black Population		0.755**	0.831***
		(0.297)	(0.264)
Share Latino Population		0.193	0.109
		(0.372)	(0.294)
Share Limited English Speaker		-0.087	0.245
		(0.984)	(0.908)
Share Foreign Born Population		-0.445	-0.513
		(0.595)	(0.524)
Share Population 65 years and older		1.005	0.876
		(0.709)	(0.611)
Standardized min. distance to a legalized state			-0.164***
			(0.054)
Having a direct flight to a legalized state			-0.008
			(0.031)
Neighboring a legalized state			0.173***
			(0.047)
Observations	245	245	245

Notes: Regressions using substate level data for the last period available (2018). Regressions include state fixed effects and standard errors are clustered at the state level.

Table A.3: Direct impact of exposure to states that have legalized marijuana on marijuana outcomes

	Panel A: Marijuana Use in Past Year			
	All Years: 2004-2020		3 Periods: 2004, 2012, 2020	
	(1)	(2)	(3)	(4)
Legalized	3.583***	3.976***	3.901***	4.406***
recreational cannabis	(1.247)	(1.209)	(1.383)	(1.403)
Substate FE	Yes	Yes	Yes	Yes
Census-Year FE	Yes	No	Yes	No
Year FE	No	Yes	No	Yes
Num of Substates	245	245	245	245
Observations	1960	1960	735	735
	Panel B: Workplace Positivity Rate for Marijuana			
	All Years: 2007-2020		3 Periods: 2009, 2012, 2020	
	(1)	(2)	(3)	(4)
Legalized	0.655***	0.664***	0.750***	0.777***
recreational cannabis	(0.149)	(0.126)	(0.159)	(0.134)
3 digit zip code FE	Yes	Yes	Yes	Yes
Census-Year FE	Yes	No	Yes	No
Year FE	No	Yes	No	Yes
Num of 3 digit zip code	753	753	753	753
Observations	10542	10542	2259	2259

Notes: It estimates the direct impact of own state's marijuana legalization on marijuana smoking behavior. Columns 1 - 2 use data from all periods, while Columns 3-4 only use data from only 3 periods: two periods prior to any marijuana legalization and one period post marijuana legalization.

Table A.4: Robustness of the impact of Out-of-State Friend Exposure to State Marijuana Legalization on Marijuana Use: heterogeneity robust estimates

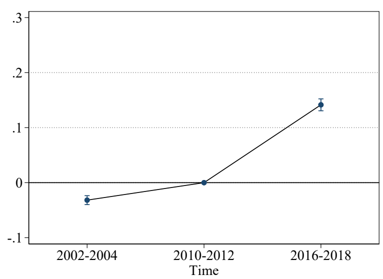
	Marijuana Use in Past Year		Workplace Positivity Rate for Marijuana	
	(1)	(2)	(3)	(4)
Log. exposure to states with legal use	1.051	1.185	0.421	0.502
	(0.539)	(0.629)	(0.132)	(0.145)
State-year FE	No	Yes	No	Yes
Treated localities	123	123	376	376
Localities	245	245	753	753
Observations	735	735	2259	2259

Notes: The coefficients are estimated using the method developed by de Chaisemartin and D'Haultfœuille (2020) and their did_multiplegt Stata package, which is available from the STATA repository. The estimator is the DID_M estimator introduced in de Chaisemartin and D'Haultfœuille (2020). It is a weighted average, across treatment values, of DID estimators comparing the change in the outcome from $t-1$ to t, in groups whose treatment changed, and in groups whose treatment remains the same. DID_M is unbiased even if the treatment effect is heterogeneous across groups. The estimator is obtained assuming that the treatment remained the same if the treatment value in the final period is less than the median. Standard errors are computed using a block bootstrap at the state level with 200 bootstrap replications. Columns 1-2 are estimated using NSDUH dataset, while Columns 3-4 are estimated using Quest.

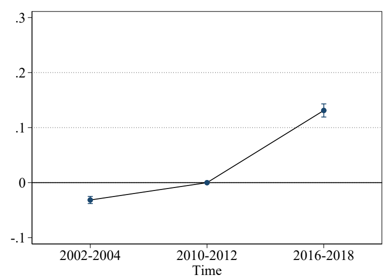
Table A.5: Impact of exposure with above and below median exposure

	Panel A: Marijuana Use in Past Year			
	All Years: 2004-2020		3 Periods: 2004, 2012, 2020	
	(1)	(2)	(3)	(4)
Above median	0.230*	0.223*	0.402**	0.406**
exposure to states with legal use	(0.126)	(0.126)	(0.169)	(0.172)
Year-state FE	Yes	Yes	Yes	Yes
Substate FE	Yes	Yes	Yes	Yes
Proximity Measures	No	Yes	No	Yes
Num of Substates	245	245	245	245
Observations	1960	1960	735	735
	Panel B: Workplace Positivity Rate for Marijuana			
	All Years: 2007-2020		3 Periods: 2009, 2012, 2020	
	(1)	(2)	(3)	(4)
Above median	0.117***	0.131***	0.231***	0.246***
exposure to states with legal use	(0.041)	(0.039)	(0.064)	(0.070)
Year-state FE	Yes	Yes	Yes	Yes
3 digit zip code FE	Yes	Yes	Yes	Yes
Proximity Measures	No	Yes	No	Yes
Num of 3 digit zip code	753	753	753	753
Observations	10542	10542	2259	2259

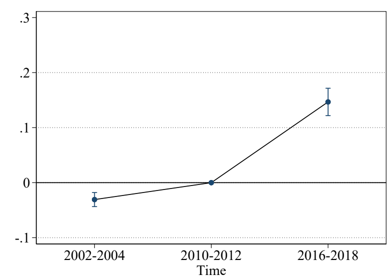
Notes: Columns 1-2 use data from all periods while Columns 3-4 only use data from only 3 periods: two periods prior to any marijuana legalization and one period post marijuana legalization. Above median exposure is defined as having above-median exposure to states with legalized use within the given year and the given state. For periods prior to any legalization, this indicator variable would be 0.



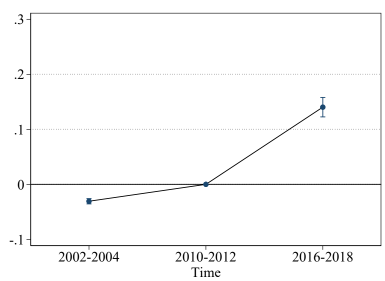
(a) Alaska



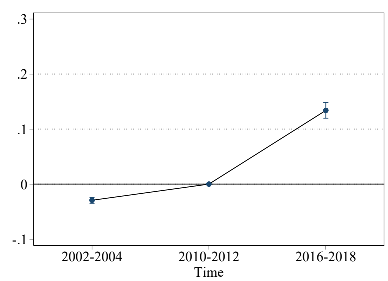
(b) Colorado



(c) Washington, D.C.



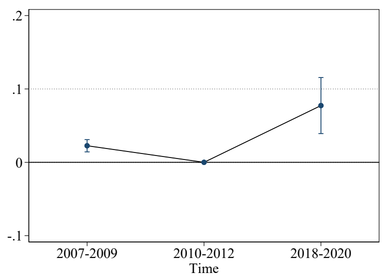
(d) Oregon



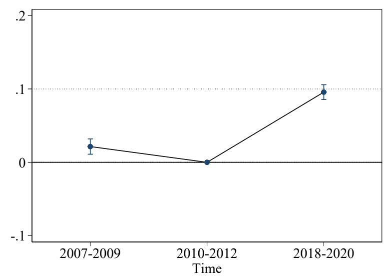
(e) Washington

Figure A.5: Event study separately by each state: the impact of the probability of having out-of-state friends exposed to a new state legalization of marijuana use on the marijuana use in the past year

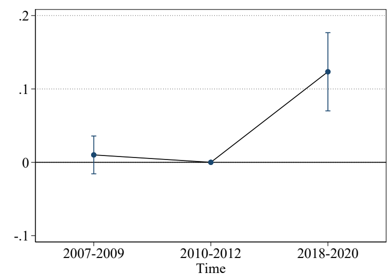
Notes: Includes only 3 time periods: 2002–2004, 2010–2012, 2016-2018. All event studies control for substate FE, state-year FE, unemployment rate, and per capita income.



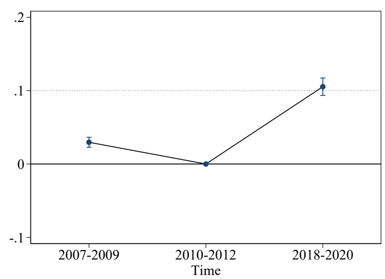
(a) Alaska



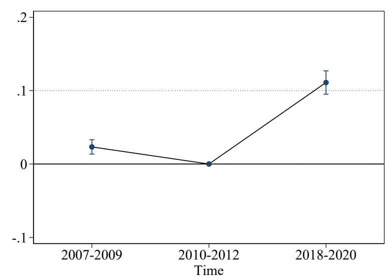
(b) Colorado



(c) Washington, D.C.



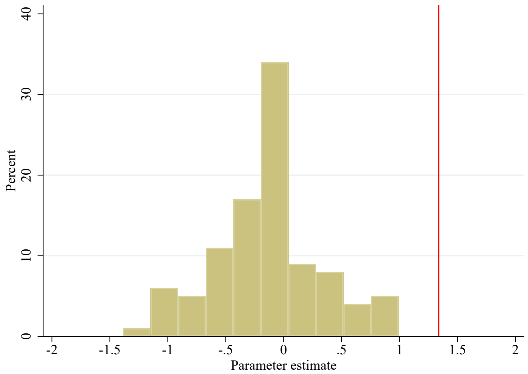
(d) Oregon



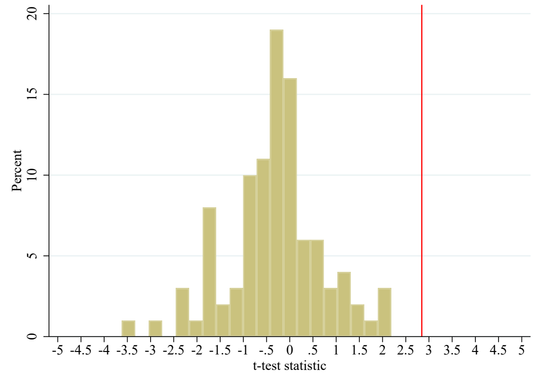
(e) Washington

Figure A.6: Event study separately by each state: the impact of the probability of having out-of-state friends exposed to a new state legalization of marijuana use on positivity rate for marijuana in workplace drug tests

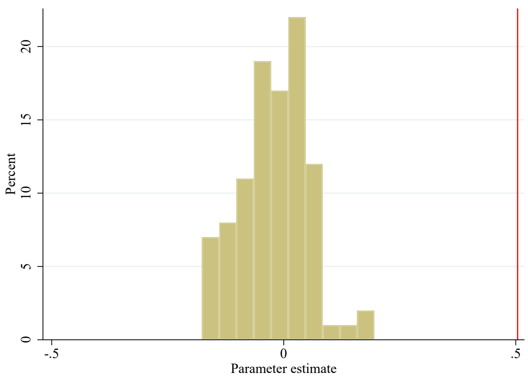
Notes: Includes only 3 time periods: 2007–2009, 2010–2012, 2018-2020. All event studies control for 3-digit-zipcode FE, state-period FE, unemployment rate and the per capita income.



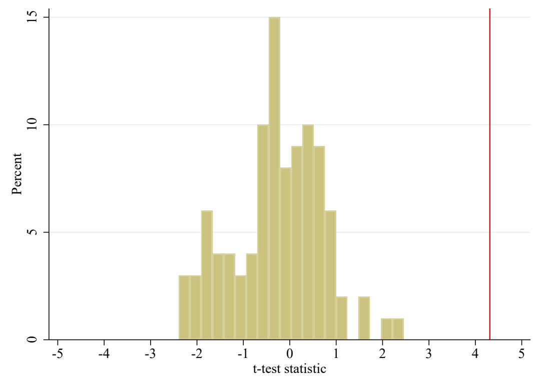
(a) Coefficients from Marijuana Use



(b) T-tests from Marijuana Use



(c) Coefficients from Workplace Drug Test



(d) T-Tests from Workplace Drug Test

Figure A.7: Histogram of Permutation Test Estimates

Notes: Each panel indicates the histogram of coefficients or t-tests from 500 permutations where the specification is the same as Equation 1, but we randomly assign network exposure within a given state and year. The red bar indicates the results from Column 6 from Table 2. Panels A.7a and A.7b utilize NSDUH while Panel A.7c and A.7d use Quest Workplace Test Dataset.

B Mechanism

To study the role of geographic characteristics, we use a three-step estimation. In the first step, to predict the social network, we regress on the logarithm of the social connectedness index between locality i and j, the logarithm of distance between the localities, and fixed effects for both localities:

$$\log(SCI_{ij}) = \alpha_1 \log(Distance_{ij}) + LocalityFE_i + LocalityFE_j + \varepsilon_{ij} \tag{6}$$

The estimates confirm what was known from the literature that distance has a high predictive power of the social network (Table B.1). Using the estimates, we obtain the predicted social connectedness index. Using the predicted index, we calculate the predicted exposure to legalization according to equation (1).

We then use a control function approach and the predicted exposure as an instrument for exposure, to estimate the following cross-sectional regression in differences:

$$\begin{aligned} MarijuanaUse_{i,T} - MarijuanaUse_{i,2012} = \beta \log[Exposure_{i,T} - Exposure_{i,2012}] \\ + StateFE_s + \varepsilon_i \end{aligned} \tag{7}$$

where the outcome variable is the change (from 2012 to the latest period T in our sample) in marijuana use in locality i in state s. The outcome variable is regressed on the logarithm of the change in exposure and state fixed effects. Note that the exposure in 2012 (before legalization) equals zero, and hence, $\log[Exposure_{i,T} - Exposure_{i,2012}] = \log Exposure_{i,T}$.

To obtain the control function estimates, first, we regress the logarithm of predicted exposure and state fixed effects on the logarithm of exposure (Table B.2):

$$\log Exposure_{i,T} = \log PredictedExposure_{i,T} + StateFE_s + \xi_i \tag{8}$$

From this regression, we obtain the residuals and include these residuals as an additional regressor in regression (7). Estimates of regression (7) without control function are presented in columns 1 and 3 of Table B.3 and with control function (residuals) in columns 2 and 4.

Table B.1: First step; outcome variable: logarithm of the social connectedness index

	Substate level (1)	3-digit zip code level (2)
α_1: Log. distance	-1.203***	-1.126***
	(0.048)	(0.065)
F-statistic of H0: $\alpha_1 = 0$	620.1	303.3
Number of states	41	48
Number of observations	71050	664899

Notes: An observation is a locality-to-locality pair. Fixed effects for both localities are included to control for the locality-specific characteristics. Standard errors are clustered at the state level.

Table B.2: Second step; outcome variable: the logarithm of exposure to legalization

	Substate level (1)	3-digit zip code level (2)
Log. predicted exposure	0.907***	0.955***
	(0.072)	(0.038)
	[0.096]	[0.075]
State FE	Yes	Yes
Number of states	41	48
Number of observations	245	753

Notes: Observation is a locality. The outcome variable is the logarithm of exposure in the latest year in the sample. The explanatory variable is the logarithm of predicted exposure in the latest year in the sample, it is predicted based on the estimates in table B.1 and calculated according to equation (1). All regressions include state fixed effects. Standard errors in parentheses are bootstrapped over steps 1–2 of the estimation with 200 bootstrap samples. Standard errors in square brackets are the naive standard errors clustered at the state level, without taking into account that the variable *Log. predicted exposure* is obtained using the estimates in step one. Stars to indicate significance are based on the bootstrapped standard errors.

Table B.3: Third step; outcome variable: change in marijuana use

	Marijuana use last year		Workplace positivity rate	
	(1)	(2)	(3)	(4)
Log. exposure	1.129**	1.306**	0.429***	0.242**
		(0.612)		(0.120)
	[0.458]	[0.594]	[0.101]	[0.106]
Residuals		-0.357		0.368**
		(0.919)		(0.176)
		[0.822]		[0.167]
State FE	Yes	Yes	Yes	Yes
Number of states	41	41	48	48
Number of observations	245	245	753	753

Notes: The table presents the control function estimates of regression (7) using the logarithm of the predicted exposure as an instrument. Observation is a locality. The outcome variable is the change (from 2012 to the latest period T in our sample) in marijuana use. The explanatory variables are the logarithm of exposure to legalization and residuals from the regression in Table B.2. All regressions include state fixed effects. Standard errors in parentheses are bootstrapped over three steps of the estimation with 200 bootstrap samples. Standard errors in square brackets are clustered at the state level, without taking into account that the variable *Residuals* is obtained using the estimates in steps 1–2. Stars to indicate significance are based on the square bracket standard errors in columns 1 and 3, and on the bootstrapped standard errors in columns 2 and 4.

C Online Appendix: Dataset construction

C.1 Facebook Social Connectedness Index

Facebook Social Connectedness Index (SCI) is available at: `https://dataforgood.facebook.com/` The measure was first constructed and analyzed by Bailey et al. (2018*b*).[10] To construct the index, Facebook users were assigned a location based on the users' information and activity on Facebook, including the stated city on their profile, and device and connection information. The 2021 public release version of the index was constructed in two steps. In step one, the index between counties i and j was calculated as:

$$SocialConnectednessIndex_{ij} = \frac{FacebookConnections_{ij}}{FacebookUsers_i \times FacebookUsers_j} \tag{9}$$

where $FacebookConnections_{ij}$ is the total number of friendship links between the two counties and $FacebookUsers_i$ is the total number of Facebook users in location i. In step two, the index was scaled to have a maximum value of one billion and a minimum of one. For privacy reasons, a small amount of random noise was added, and locations with too few number of users were excluded. The index measures the relative probability of a Facebook friendship link between a given Facebook user in location i and a given Facebook user in location j. If the index is twice as large, a given Facebook user in location i is about twice as likely to be friends with a given Facebook user in location j. The 2021 public release version of the index was calculated as of October 2021.

C.2 Aggregation of the SCI

The finest geographical levels available for the 2021 SCI are US ZCTA (zip code tabulation) and county. Our outcome variables are at the substate level (aggregation of counties) for the NSDUH data and 3 digit-zip code level (aggregation of zip codes) for the Drug Testing Index from Quest Diagnostic. Thus, we need to construct measures of the SCI at the substate and 3 digit-zip code level to match the level of our outcome variables. To do this, we follow the aggregation instructions in the Social Connectedness Index Data Notes.[11]

For the aggregation of the county level SCI to the substate level SCI, let us index the counties in each substate i by $r_i \in R(i)$, let $Friendships_{ri,rj}$ count the total number of friendship links between individuals in counties r_i and r_j, let Pop_{ri}, denote the total (Facebook) population in county r_i, and denote $PopShare_{ri}$, denote the share of that population

[10] The terminology has changed over time: the measure that is called the Social Connectedness Index in the public release data of 2021, was called the relative probability of friendship by Bailey et al. (2018*b*).

[11] Source: `https://dataforgood.facebook.com/dfg/docs/methodology-social-connectedness-index`

in county r_i in substate i: $\sum_{ri \in R(i)} PopShare_{ri} = 1$. Then the SCI between substate i and substate j is given by:

$$\begin{aligned} SCI_{i,j} &= \frac{Friendships_{i,j}}{Pop_i \times Pop_j} = \frac{\sum_{ri \in R(i)} \sum_{rj \in R(j)} Friendships_{ri,rj}}{\left(\sum_{ri \in R(i)} Pop_{ri}\right) \times \left(\sum_{rj \in R(i)} Pop_{rj}\right)} \\ &= \sum_{ri \in R(i)} \sum_{rj \in R(j)} \frac{Pop_{r_i}}{\sum_{ri \in R(i)} Pop_{ri}} \frac{Pop_{rj}}{\sum_{rj \in R(j)} Pop_{rj}} \frac{Friendships_{ri,rj}}{Pop_{ri} \times Pop_{rj}} \\ &= \sum_{ri \in R(i)} \sum_{rj \in R(j)} PopShare_{r_i} \times PopShare_{r_j} \times SCI_{ri,rj} \end{aligned} \tag{10}$$

Similarly, we use equation (10) to aggregate the ZCTA level SCI to the 3 digit-zip code level SCI. We assume that zip code and ZCTA are the same.

C.3 Construction of locality level unemployment and income measures

We use the county-level data to construct substate and 3-digit zip code unemployment and income measures. While the majority of substates are defined in terms of counties, some are defined in terms of census tracts. For these substates defined by census tracts, we use income and unemployment rate for the largest county (measured by population) in each substate. To calculate the 3-digit zip code measures, we match each ZIP Code Tabulation Area (ZCTA) with the county with the most population overlapped.