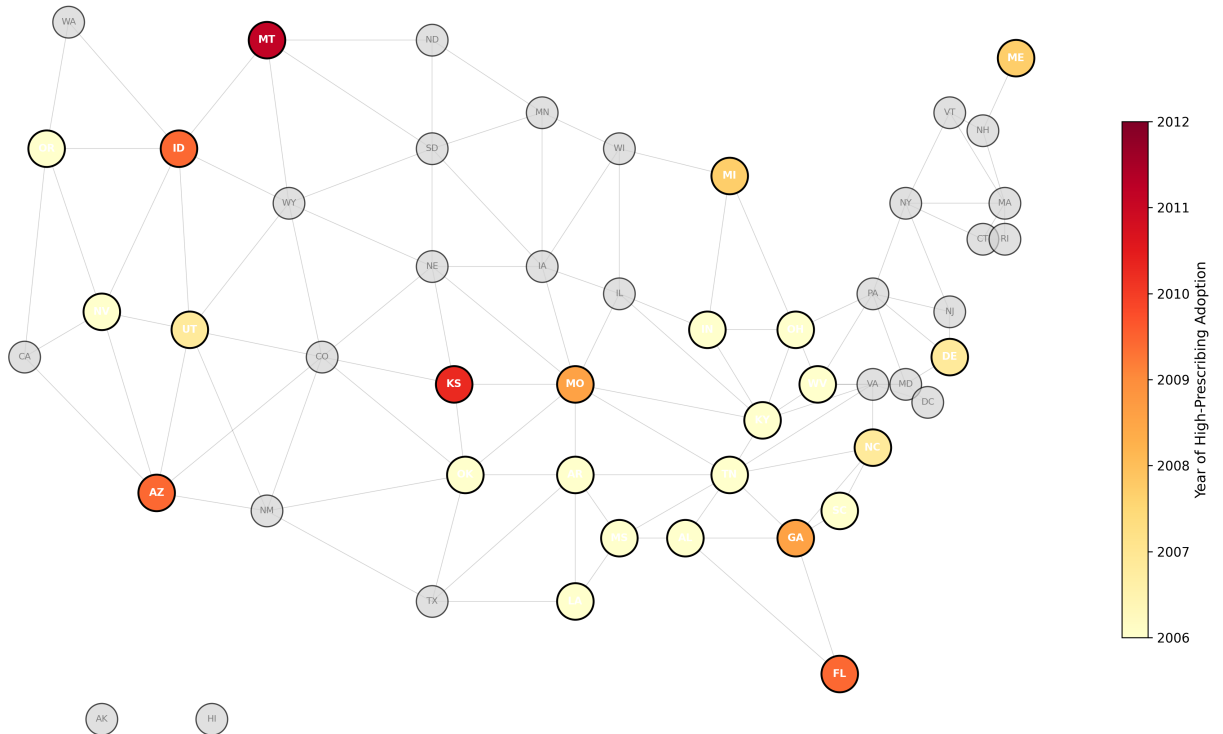


Opioid Epidemic Network Modeling

Vishal Maradana
Georgia Institute of Technology
Atlanta, GA, United States
vmaradana3@gatech.edu

Rishi Bandi
Georgia Institute of Technology
Atlanta, GA, United States
rbandi6@gatech.edu

**Geographic Distribution of High-Prescribing Adoption
(Gray = Never Exceeded Threshold)**



Abstract

This project models high-rate opioid prescribing as a contagious behavior diffusing across U.S. states through geographic and professional proximity. Using CDC dispensing data (2006–2018), we infer a directed influence network among states and apply spatial autoregressive models to predict prescribing rates. Our state-level model achieves $R^2 = 0.884$, with 98.1% self-persistence and a small but significant 0.9% neighbor spillover effect. Tennessee, Nevada, and South Carolina emerge as top influencers. At the county level, we identify 345 “superspreader” counties that adopted high prescribing before their states, with Texas counties dominating intra-state influence rankings. The county model ($R^2 = 0.792$) reveals stronger spatial effects (5%) than the state model, suggesting finer-scale

geographic dynamics. These findings support regionally targeted interventions over state-by-state approaches.

1 Introduction

The opioid crisis continues to wreck communities across America. Starting in the late '90s, overdoses tied to these drugs have killed more than half a million people - prescription painkillers helped kick off this wave. Experts often look at doctors' choices, rules for treatment, or traits in patients, yet pay little mind to how those habits jump from place to place. New findings show healthcare actions aren't born out of thin air; instead, they travel via work circles, hospital cultures, local trends, and laws nearby. That leads to a key thought: might heavy prescribing act like a bug spreading between towns and regions?

Seeing prescription habits as something shaped by location changes how we plan solutions. Because when one area shows

rising trends before nearby spots do, tackling those starting zones might slow wider spread. But if each place’s use mostly depends on its own conditions - and doesn’t spill across borders - then fixes need to match local realities instead. Since causes differ, tools must spot geographic links clearly while showing just how much they actually predict change.

This research looks at opioid prescriptions like something that might spread from place to place, checking if high usage moves between neighboring states and counties. With help from CDC numbers covering 2006 up to 2018, we tackle two linked goals. One, we map out one-way influence paths - statewide and county-wide - to spot areas that led the shift toward heavy prescribing before others followed. Instead of just adding points together, we follow shifts each year into “high” use zones, guess how places may have affected one another, then sort them by importance in this web of connections. Next, we build models that guess coming prescription levels by looking at a place’s past numbers along with nearby areas’ averages. Using this method helps split trends into two parts: one driven by local history, the other shaped by what’s happening around it.

Running checks on both state and county data helps us see if location-based trends shift across bigger or smaller regions. Because results might differ by scale, this look can show where prevention efforts work best - be it broad zones, standout states, or particular counties spreading unsafe prescription habits ahead of nearby spots.

2 Related Work

Work on diffusion and influence in networks provides the algorithmic backbone for our project. Kempe, Kleinberg, and Tardos [3] formalized two canonical views of spread—the Independent Cascade (IC) and Linear Threshold (LT) models—demonstrating that influence maximization is submodular and admits a tight greedy approximation. Their framework clarifies why heuristic targeting (e.g., high degree) often fails by ignoring redundancy. However, their approach assumes a known network and focuses on forward optimization (who to seed), whereas our challenge is the inverse problem: inferring the underlying influence structure from observed adoption times.

To address this inverse problem, Gomez-Rodriguez et al. developed NETINF [1], which infers directed edges from cascade timing data using a submodular optimization approach. While NETINF effectively recovers network structures in large-scale web data, its reliance on tree-structured cascades and multiple independent contagion events does not align perfectly with the single, cyclical annual stream of state-level opioid data. Consequently, we adopt their timing-to-edges intuition—using adoption order to reveal hidden edges—but employ a transparent, count-based construction method better suited for a single macroscopic trajectory.

In the domain of public health networks, Kaminski et al. [2] constructed a provider network in Indiana based on shared patients, revealing that data-driven clinical communities often diverge from administrative districts. This highlights the limitations of purely geographic targeting and suggests that influence flows through professional relationships. Similarly, O’Malley et al. [4] quantified risky prescribing within shared-patient networks, finding that homophily exists at the supra-dyadic (triadic) level. Their findings suggest that

local norms and clustered structures help sustain behavior, cautioning against interventions that focus solely on single “hubs.” However, both studies focus on clinician-level snapshots without explicit counterfactual policy simulations. Our work scales these insights to the state level to analyze inter-jurisdictional influence over nearly two decades.

Recent work by Yang et al. [5] further expanded this field by demonstrating that bipartite-aware centrality measures outperform traditional algorithms in detecting high-risk prescribing patterns. Collectively, this literature suggests a coherent path: we leverage the diffusion logic of [3] and the inference perspective of [1], but adapt them to the constraints of aggregate health data. By bridging the structural insights of [2] and [4] with policy timelines, we address a critical gap: linking network structures to policy-timed counterfactuals at the jurisdictional scale. While previous works implicitly caution against overclaiming causality, we position our inferred influence edges as directional hypotheses validated through robustness checks and policy simulations.

3 Problem Definition

We investigate whether high opioid prescribing rates spread geographically across U.S. states and counties, or whether they emerge independently in each region. Specifically, we address two questions. First, do regions that become high-prescribing tend to be geographically adjacent to regions that were already high-prescribing? If so, this suggests a spatial diffusion process where prescribing norms spread across boundaries. Second, can we identify specific states or counties that consistently preceded others in adopting high prescribing behavior, acting as “influencers” or early adopters that may have driven broader regional trends? Answering these questions requires distinguishing between two scenarios: one where prescribing patterns are driven primarily by local factors (demographics, healthcare access, economic conditions) with minimal cross-boundary spillover, and another where a region’s prescribing behavior is meaningfully shaped by what happens in neighboring regions. Understanding which scenario better describes the data has direct implications for intervention design, whether to target isolated high-risk areas or to address regional clusters through coordinated multi-state efforts.

4 Data Collection and Preprocessing

Opioid dispensing data came from the CDC’s public resources. Historical state-level data (2006–2018) were collected by scraping archived pages from the CDC’s drug overdose prevention website. For each year, we extracted state names, abbreviations, and rates per 100 persons, mapping abbreviations to FIPS codes using a lookup table. More recent state-level data (2019–2023) came directly from the CDC’s Opioid Dispensing Rate Maps page.

County-level data required separate preprocessing. We parsed archived text files (2006–2018) containing county dispensing rates, handling multiple file formats where county-state ordering varied across years. Each record was standardized to include county name, state abbreviation, FIPS code, and dispensing rate. County names were normalized with proper suffixes (County, Parish, Borough) and matched to state identifiers. Missing values (marked as dashes or “Data unavailable”) were preserved for later handling. Recent

county data (2019–2023) were downloaded directly from the CDC, covering approximately 98% of U.S. counties.

Preprocessing merged historical and recent panels for both geographic levels, standardizing columns and removing national aggregate rows. Text-formatted rates were converted to numeric values. Duplicate state-year and county-year entries were resolved by keeping the most recent record. The final datasets :a state panel (2006–2018) and county panel (2006–2018).

5 Methods

Our approach combines two complementary techniques: (1) influence network construction to identify which regions preceded others in adopting high-prescribing behavior, and (2) spatial autoregressive modeling to predict future prescribing rates and quantify spatial spillover effects. We apply both methods at state and county levels to examine whether spatial dynamics differ across geographic scales.

5.1 Defining High-Prescribing Status

We initially divide regions into those that are classified as high-prescribing and those that are not, for each year, when investigating the prescribing rates. For our state-level analysis, we use the 75th percentile of the national distribution as a reference point, 87.35 prescriptions per 100 persons. A state in a given year is considered to be high-prescribing if its prescribing rate surpasses this threshold.

Our unique threshold-based method converts the continuous prescribing rates of each state into a simple, binary system, and allows us to draw inferences within a network. We find it necessary to consider county-level analysis, so we apply regional state-specific cutoffs that differ from a single nationwide threshold.

In doing so, we determine within each state the 75th percentile of county prescribing rates across all years, and reclassify counties relative to their own state’s distribution.

5.2 Influence Network Construction

Think about it this way. Most old-school studies on opioid prescriptions look at places in isolation - like checking only what’s inside one town without peeking next door. But here’s the thing - what happens in one area might actually nudge another nearby place into doing the same. Instead of ignoring that ripple effect, our method uses connections between spots over time. Here’s how - it checks if spot A had heavy prescribing just before spot B started ramping up. That timing clue hints that A could’ve shaped B’s habits. Tally up those patterns year after year, you get a map with arrows pointing from likely influencers to followers - with thicker lines showing stronger links.

Step by step. For every two years in a row ($t, t+1$), we check what’s happening

- **Starting group** S_t : areas with heavy use in year t , yet showing strong trends back then
- **New adopter set** N_{t+1} : areas prescribing a lot in year $t + 1$, yet didn’t do so in year t

From each source $s \in S_t$ to every fresh adopter $n \in N_{t+1}$, draw an arrow $s \rightarrow n$. These links build up over time - but older ones fade slowly, so lasting early impact counts more

$$w(s \rightarrow n) = \sum_{t:s \in S_t, n \in N_{t+1}} \frac{1}{1 + (t - t_s)} \quad (1)$$

where t_s stands for the year region s started adopting. This method gives more weight to areas that began early but also kept prescribing at high levels over time - since those who stuck with it likely shaped peers more than those who joined later or only briefly.

State-Level Network. For each state, connections only go to nearby states - ones they actually touch. A list of neighbors is made using common borders; like how Georgia links to Alabama, Florida, North Carolina, South Carolina, or Tennessee. Being close matters because health systems often overlap across these lines, so ideas and practices spread easier from one side to the next.

County-Level Network. In every state, we set up unique influence maps for each county. Instead of mixing areas across states, this method tracks how heavy prescribing moves between nearby counties under one government. It skips messy inter-state links but shows clear local trends.

5.3 Ranking Influential Regions

After building the influence map, folks are ranked based on how likely they are to pass along prescription habits. Their position is measured using a score that counts outgoing links with added weight

$$C_{out}(s) = \sum_{n \in V} w(s \rightarrow n) \quad (2)$$

where V stands for all nodes, while $w(s \rightarrow n)$ means the link strength from s to n - set to zero when there’s no connection. A higher out-degree shows one area came before many others, carrying heavier impact, meaning it likely shaped more choices.

We calculate eigenvector centrality too, since it considers how impactful a node’s connections are. Instead of just counting links, being tied to key areas boosts your score more than linking to minor ones. Influence spreads step by step - so when hubs link together, their impact multiplies down the line.

5.4 Spatial Autoregressive Model

Gut feeling says the network shows past trends, yet won’t tell you how much nearby areas affect what’s coming next. So instead, we built a SAR setup - one that splits local habits from outside influences. When outside effects don’t matter, each place just follows its own path. But when they do, neighbors add clues all on their own.

Model Details. In every area s and time t , we predict next year’s prescription level by:

$$R_{s,t+1} = \alpha + \beta \cdot R_{s,t} + \gamma \cdot \bar{R}_{N(s),t} + \epsilon_{s,t} \quad (3)$$

where:

- $R_{s,t}$ is the current prescribing rate (self-history feature)
- $\bar{R}_{N(s),t}$ is the average rate of s 's geographic neighbors (spatial feature)
- β is the self-persistence coefficient
- γ is the spatial spillover coefficient
- $\epsilon_{s,t}$ is the error term

At state level, nearby areas count if they share a border. If a state stands alone - like Alaska or Hawaii - it gets no neighbor effect. The setup uses basic linear math to learn patterns over time. Data from 2006 to 2016 trains the system; what happens in 2017 and 2018 checks how well it works.

At the county level, implementation uses each state's overall average instead of nearby counties' numbers. That shows whether a local prescribing pattern leans more on statewide trends - using that as a backdrop

$$R_{c,t+1} = \alpha + \beta \cdot R_{c,t} + \gamma \cdot R_{state(c),t} + \epsilon_{c,t} \quad (4)$$

where $R_{state(c),t}$ stands for the rate at that time in the state where county c sits. This setup checks if counties react to broader state shifts, not just their past patterns.

The coefficient β shows how much a place sticks to its own past - like whether old trends shape what happens next. Near-1 values mean things shift gradually, locked into history. Meanwhile, captures outside influence - how much nearby areas or broader patterns add to local trends. If γ is above zero, it hints that habits like prescribing spread across regions. When γ hovers near zero, actions seem more tied to internal conditions than neighbors.

5.5 Superspreader Identification

In local areas, some spots started heavy prescribing early - way ahead of their state's average. Think places where:

$$t_{adopt}^{county} < t_{adopt}^{state} \quad (5)$$

Some places saw spikes before the rest of their states did. Spotting those spots shows where heavy prescribing started, also highlights where quick action could help most.

5.6 Evaluation Metrics

We check the space-based model with usual fit measures on the kept-back test data:

- **R^2 Score:** How much of the changes in upcoming rates the model actually accounts for
- **Mean Squared Error (MSE):** Average squared prediction error

For influence rankings, we list the top-k key areas, then check if results stay similar using different methods - like out-degree or eigenvector. When both approaches agree, it's more likely these hubs aren't just flukes from one technique but actually matter.

6 Results

6.1 Experimental Questions and Testbed

Our experiments address three core questions:

- (1) **Q1:** Can we predict future opioid prescribing rates using spatial autoregressive models that incorporate neighbor effects?
- (2) **Q2:** Which states and counties act as "influencers" that preceded others in adopting high-prescribing behavior?
- (3) **Q3:** Do spatial dynamics differ between state-level and county-level analyses?

Testbed. We use CDC opioid dispensing data (2006–2018) comprising 51 states and ~3,000 counties per year. The high-prescribing threshold is the 75th percentile (87.35 per 100 persons for states; state-specific thresholds for counties). Models are trained on 2006–2016 and tested on 2017–2018.

6.2 State-Level Prediction Model (Q1)

The spatial autoregressive model achieves $R^2 = 0.884$ on the held-out test set. Table 1 presents the estimated coefficients.

Table 1: State-Level Spatial Autoregressive Model

Parameter	Estimate	Interpretation
β (Self-Persistence)	0.981	98.1% persistence
γ (Spatial Spillover)	0.009	~1% neighbor effect
R^2	0.884	
MSE	21.58	

The high self-persistence ($\beta = 0.981$) indicates prescribing rates are path-dependent—states that reach high rates tend to stay there. The spatial spillover ($\gamma = 0.009$), while small, confirms that neighbor rates carry predictive information beyond self-history. Figure 1 shows actual vs. predicted rates.

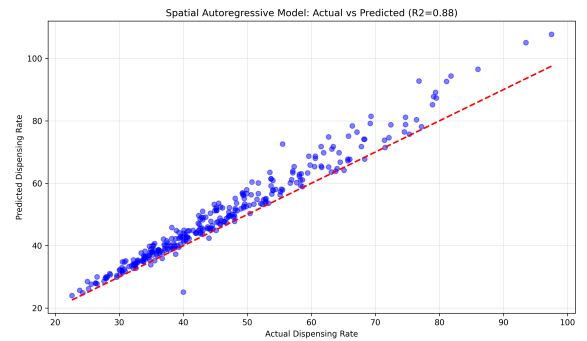


Figure 1: State-level prediction: Actual vs. predicted rates ($R^2 = 0.884$).

6.3 State-Level Influence Network (Q2)

The influence network contains 20 nodes and 21 directed edges. Table 2 shows the top influential states by weighted out-degree centrality.

Table 2: Top 5 Influential States

Rank	State	Out-Degree Weight
1	Tennessee (TN)	1.67
2	Nevada (NV)	1.50
3	South Carolina (SC)	1.33
4	Georgia (GA)	1.00
5	Alabama (AL)	0.58

Tennessee emerges as the primary influencer, consistent with its early adoption (2006), central location, and sustained high rates. Key influence pathways include GA→FL, NV→UT, and TN→NC.

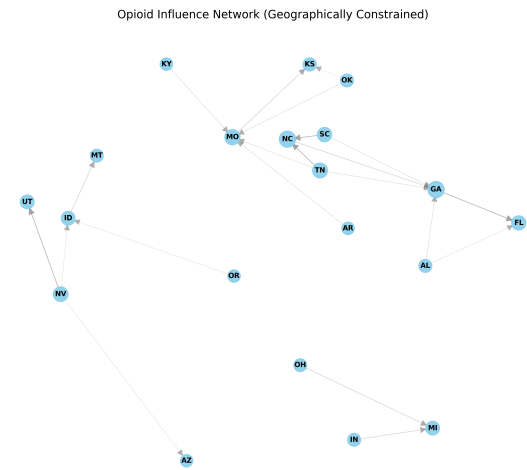


Figure 2: State-level influence network. Node size reflects out-degree; Southern states form a densely connected cluster.

6.4 County-Level Results (Q3)

6.4.1 *County Prediction Model.* The county-level model uses state average rates as the spatial feature. Table 3 compares the national and Georgia-specific models.

Table 3: County-Level Model Comparison

Metric	All Counties	Georgia Only
β (Self-Persistence)	0.950	0.968
γ (State Effect)	0.050	0.315
R^2	0.792	0.783

County-level shows stronger spatial effects ($\gamma = 0.050$) than state-level ($\gamma = 0.009$), indicating finer-scale geographic dynamics. The Georgia model shows even higher state influence ($\gamma = 0.315$), suggesting state trends strongly affect county behavior in that region. Figure 3 shows the county-level prediction performance.

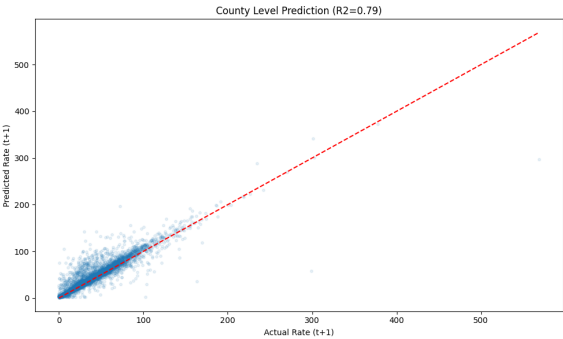


Figure 3: County-level prediction: Actual vs. predicted rates ($R^2 = 0.792$).

6.4.2 *Superspreader Counties.* We identified 345 counties that adopted high rates before their state crossed the threshold. Figure 4 shows the top superspreader county from each state. Montana counties led by 6 years, followed by Kansas (5 years) and Idaho (4 years). These early “hotspots” represent potential targets for early intervention.

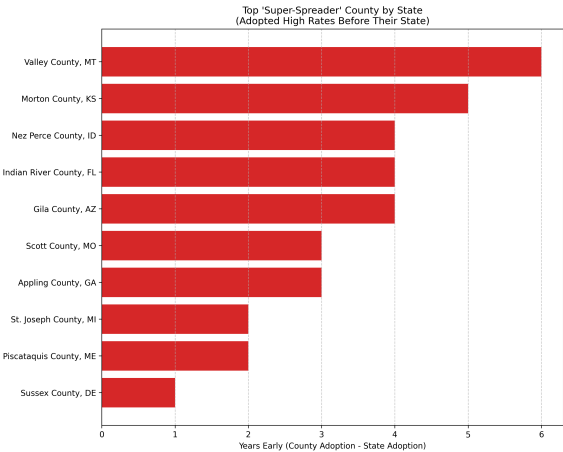


Figure 4: Top superspreader counties by state, showing years adopted before state threshold.

6.5 Summary of Findings

- (1) **Strong Prediction Performance:** State-level model achieves $R^2 = 0.884$; county-level achieves $R^2 = 0.792$.
- (2) **High Path Dependence:** Self-persistence dominates ($\beta > 0.95$), meaning regions that become high-prescribers tend to remain so.
- (3) **Measurable Spatial Spillover:** Neighbor effects are small but detectable, stronger at county level (5%) than state level (1%).
- (4) **Key Influencers:** Tennessee, Nevada, and South Carolina are top state-level influencers; 345 superspreader counties preceded state-level trends.

7 Conclusion

This study modeled opioid prescribing as a spatial contagion process, examining whether high-rate prescribing diffuses across U.S. states and counties. Our spatial autoregressive models achieved strong predictive performance ($R^2 = 0.884$ at state level, $R^2 = 0.792$ at county level), demonstrating that future prescribing rates can be reliably predicted from historical patterns and geographic context. The high self-persistence coefficients ($\beta > 0.95$) confirm that prescribing behavior is highly path-dependent—regions that become high-prescribers tend to remain so. While spatial spillover effects are modest (1% at state level, 5% at county level), they provide measurable predictive value beyond self-history alone.

Our influence network analysis identified Tennessee, Nevada, and South Carolina as key state-level influencers that consistently preceded neighboring states in adopting high-prescribing behavior. At the county level, we discovered 345 “superspreader” counties that adopted high rates years before their states crossed the threshold, with Montana counties leading by 6 years. These findings suggest that early, localized hotspots may seed broader regional epidemics.

The results support regionally coordinated interventions over isolated state-by-state approaches. Given the strong path dependence observed, prevention of initial high-rate adoption is likely more effective than attempting to reverse established patterns.

8 Future Work

Future research will focus on two key extensions. First, we plan to incorporate **policy intervention simulations** that model the

counterfactual impact of targeting high-influence states or superspreader counties. By simulating early interventions in identified influencer regions, we can estimate how many downstream adoptions might have been prevented. Second, we aim to develop **preventive measure modeling** that evaluates strategies such as Prescription Drug Monitoring Program (PDMP) implementation timing, prescriber education campaigns, and cross-state coordination efforts. These simulations would help public health authorities allocate intervention resources more effectively by quantifying the network-wide benefits of targeting specific regions versus broad-based approaches.

References

- [1] Manuel Gomez-Rodriguez, Jure Leskovec, and Bernhard Schölkopf. 2010. Inferring Networks of Diffusion and Influence. In *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD '10)*. ACM, Washington, DC. doi:10.1145/1835804.1835933
- [2] Patricia M. Kaminski, Thomas E. Kottke, Chiahui Yen, Claudia B. Holt, Tyler N. A. Winkelman, Lacey Hartman, Ruby Nguyen, and Molly M. McCarthy. 2023. Comparing professional communities: Opioid prescriber networks and Public Health Preparedness Districts. *Harm Reduction Journal* 20, 1 (2023), 120. doi:10.1186/s12954-023-00840-8
- [3] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the Spread of Influence through a Social Network. In *Proc. Ninth ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD '03)*. ACM, Washington, DC. doi:10.1145/956750.956769
- [4] Xin Ran, Ellen Meara, Nancy E. Morden, Erika L. Moen, Daniel N. Rockmore, and A. James O'Malley. 2024. Estimating the impact of physician risky-prescribing on the network structure underlying physician shared-patient relationships. *Applied Network Science* 9 (2024), Article 63. doi:10.1007/s41109-024-00670-y
- [5] Jie Yang, Qian Li, Aaron J. Smith, and Kimberly W. Johnson. 2022. Comparing Centrality Measures in Bipartite Patient–Provider Networks for Identifying High-Risk Opioid Prescribing and Overdose. *J. Biomed. Inform.* 136 (2022), 104234. doi:10.1016/j.jbi.2022.104234

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009