Applied Network Science

RESEARCH Open Access

Estimating the impact of physician risky-prescribing on the network structure underlying physician shared-patient relationships

Xin Ran[1,2], Ellen Meara[3,4], Nancy E. Morden[2,5], Erika L. Moen[1,2], Daniel N. Rockmore[6,7,8] and A. James O'Malley[1,2,6,7*]

*Correspondence:
james.omalley@dartmouth.edu

[1] Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA
[2] The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA
[3] Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[4] National Bureau of Economic Research, Cambridge, MA 02139, USA
[5] United HealthCare, Minnetonka, MN 55343, USA
[6] Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA
[7] Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA
[8] The Santa Fe Institute, Santa Fe, NM 87502, USA

Abstract

Social network analysis and shared-patient physician networks have become effective ways of studying physician collaborations. Assortative mixing or "homophily" is the network phenomenon whereby the propensity for similar individuals to form ties is greater than for dissimilar individuals. Motivated by the public health concern of risky-prescribing among older patients in the United States, we develop network models and tests involving novel network measures to study whether there is evidence of homophily in prescribing and deprescribing in the specific shared-patient network of physicians linked to the US state of Ohio in 2014. Evidence of homophily in risky-prescribing would imply that prescribing behaviors help shape physician networks and would suggest strategies for interventions seeking to reduce risky-prescribing (e.g., strategies to directly reduce risky prescribing might be most effective if applied as group interventions to risky prescribing physicians connected through the network and the connections between these physicians could be targeted by tie dissolution interventions as an indirect way of reducing risky prescribing). Furthermore, if such effects varied depending on the structural features of a physician's position in the network (e.g., by whether or not they are involved in cliques—groups of actors that are fully connected to each other—such as closed triangles in the case of three actors), this would further strengthen the case for targeting groups of physicians involved in risky prescribing and the network connections between them for interventions. Using accompanying Medicare Part D data, we converted patient longitudinal prescription receipts into novel measures of the intensity of each physician's risky-prescribing. Exponential random graph models were used to simultaneously estimate the importance of homophily in prescribing and deprescribing in the network beyond the characteristics of physician specialty (or other metadata) and network-derived features. In addition, novel network measures were introduced to allow homophily to be characterized in relation to specific triadic (three-actor) structural configurations in the network with associated non-parametric randomization tests to evaluate their statistical significance in the network against the null hypothesis of no such phenomena. We found physician homophily in prescribing and deprescribing. We also found that physicians

exhibited within-triad homophily in risky-prescribing, with the prevalence of homophilic triads significantly higher than expected by chance absent homophily. These results may explain why communities of prescribers emerge and evolve, helping to justify group-level prescriber interventions. The methodology may be applied, adapted or generalized to study homophily and its generalizations on other network and attribute combinations involving analogous shared-patient networks and more generally using other kinds of network data underlying other kinds of social phenomena.

Keywords: Risky prescribing, Shared-patient physician network, Homophily, Deprescribing, Quantifying polypharmacy, State-space, Transition matrix

Introduction

The excessive prescribing ("polypharmacy") of unwarranted prescriptions that deviate from guidelines (Dreischulte and Guthrie 2012; Gnjidic et al. 2013; Bushardt et al. 2008) commonly known as risky prescribing is a health concern, particularly among the older population, for which public health interventions are critically needed. In the U.S., the older population consumes more than one-third of prescription medications, yet they consist of around 15% of the population (Fulton and Riley Allen 2005; Werder and Preskorn 2003). Even more concerning are the adverse events associated with risky prescribing. For example, evidence was found that the combined usage of opioids and benzodiazepines or non-benzodiazepine sedative-hypnotics (sedative-hypnotics) has a higher risk of overdose than using opioids alone (Cho et al. 2020; Centers for Disease Control and Prevention and others 2016; Sun et al. 2017).

Physicians are a major determinant of patient drug regimes, especially for drugs that patients cannot directly access. Social network analysis has proven to be effective for studying collaborations among physicians and their association with patients' health outcomes (Barnett et al. 2012; Fattore et al. 2009; Pollack et al. 2015; Moen et al. 2016). Understanding how different prescribing behaviors are embedded in the shared-patient physician network may help identify candidate physicians to intervene on in order for the impact of the intervention to be maximized. This would be the case if the most connected physicians also transmitted the effects of a behavior change intervention spillover to the greatest number of other physicians; e.g., through a process of peer-effects (Ran et al. 2024b). Similarly, if actors with certain shared or similar traits are clustered together in the network, targeting groups of connected persons exhibiting undesirable traits might be the most effective way of implementing a behavior change intervention. Assortative mixing or "homophily", commonly known as "birds of a feather flock together", is a social phenomenon in which people who share similar traits are more likely to form relationships with each other (McPherson et al. 2001; Apicella et al. 2012; Rand et al. 2011). Because individuals are more prone to interact with individuals they resemble than those they don't, the existence of homophily can reinforce shared behaviors between pairs ("dyads") or larger groups of connected individuals in social networks (Centola 2011). Therefore, knowing that homophily exists suggests that interventions that positively change the behavior of individuals might also seek to dissolve problematic ties (e.g., those in which both actors are risky prescribers).

Previous studies have found that physicians with the same organizational affiliation were more likely to develop professional relationships (Landon et al. 2012; Mascia et al.

2015) and that homophily in a network of opioid users was associated with the number, type and daily dosage of opioid prescriptions (Aroke et al. 2021). These prior works motivate the conjecture that homophily on risky prescribing may generalize to more complex phenomena than a dyadic phenomenon resulting in the emergence of clusters of three or more heavy prescribers. However, with the exception of the preliminary work by Ran et al. (2024b), to our knowledge structural characteristics of shared-patient physician networks of prescribers of risky drugs such as those in the opioids, benzodiazepines and sedative-hypnotics classes is understudied.

Several studies have quantified patients' receipt of polypharmacy and physicians' opioid prescribing patterns among different specialties (Quinn and Shah 2017; Levy et al. 2015). However, current approaches for quantifying physicians' prescribing behaviors identify risky prescribing without accounting for the extent to which appropriate deprescribing occurs. Unlike prescribing, deprescribing often takes place in conversations during physician-patient encounters involving reviews of patients' medications (Farrell and Mangin 2019) and triggers no insurance claim. This leads to challenges in identifying the physician or physicians responsible for deprescribing. An important contribution within the emerging field of data science is the development of heuristic algorithms for identifying likely instances of deprescribing and the physicians responsible in claims data.

Exponential random graph models (ERGMs) provide a general modeling framework for relating network phenomena and actor attributes to the likelihood of observing a network. In theory, they provide an ideal methodological basis for estimating which factors, including those relating to physician prescribing behavior, are most strongly associated with the observed network. One challenge with ERGMs is the phenomenon in which the model puts most of its mass on a very dense or sparse network. Known as degeneracy, this phenomenon has been commonly encountered by investigators examining whether dyadic-dependent network phenomena such as transitivity underlie the network (Handcock et al. 2008; Handcock 2003; Moen et al. 2016). When using ERGMs to study homophily, degeneracy may limit our ability to isolate the true level of homophily from the confounding effects of other network phenomena.

To avoid degeneracy while still testing for the presence of higher-order (extra-dyadic) homophily, we introduce two new network statistics that capture specific triadic phenomena and develop statistical tests of whether their prevalence in the network exceeds that expected by chance in the absence of homophily. The two new statistics quantify risky-prescribing associated homophily acting within closed triadic configurations ("triangles"). To evaluate whether the observed prevalence of such configurations in the network is statistically significant, we construct a permutation test that randomly redistributes the node attributes across the nodes in the network to evaluate the null distribution of the test statistic. Using these new statistics and associated statistical tests, we studied the extent to which risky prescribing is associated with the closed triads of physician relationships.

The specific contributions of our work include:

- Investigation of prescribing-associated homophily in physician shared-patient networks using ERGMs adjusting for other network statistics and the prevalence of node

attributes. In addition, we derive identifiability conditions for the effect of homophily when including actor-specific node-attributes of the same variables defining the homophily measures in order to isolate the pure effect of homophily.
- Development of novel triadic network statistics and non-parametric statistical tests for determining whether homophily generalizes to exhibit an extra-dyadic effect (i.e., acts beyond dyads) while avoiding degeneracy.
- A novel framework to quantify physician prescribing behavior by their contribution to patient prescription state transitions and the development of multiple prescribing indexes that can be conveniently computed and used as node attributes to study homophily of physician prescribing.

Motivating application: risky prescribing study

The methodological research in this paper is motivated by whether physicians who engage in prescribing of three classes of risky drugs—opioids, benzodiazepines and sedative-hypnotics—to US Medicare beneficiaries are more likely to be connected in a physician professional relationship network than physicians whose prescribing tendencies differ from one another. We hypothesis that this dependency extends beyond dyads to a higher-order network phenomena (i.e., configurations of three or more physicians). The dependent variables are the physician network and quantities derived from it while the key predictors are physician prescribing indexes constructed from sequences of patient prescription drug states. Distinct data sets and data wrangling pipelines (see Figure 1 in the Supplemental Appendix) were used to construct the physician network and the prescribing indexes for each physician. These are outlined in the remainder of this section and described in more depth in the companion methodological sections on modeling the network data and constructing the prescribing indexes in sections "Network Modeling and statistical analysis methods" and "Measures of homophily in physician prescribing and deprescribing", respectively.

Overview of physician network formation and methods

We used a 40% random sample of all Medicare fee-for-service claims in 2014 of beneficiaries residing in the state of Ohio (the only state for which we had patient-physician encounter data) in 2014 to extract a dataset of physician-patient encounters. Initially, a directed physician network was constructed by assuming that a visit to physician i followed by a visit to physician j by the same patient within 2014 provides evidence of a meaningful professional relationship (e.g., a genuine "patient referral") from physician i to j (O'Malley et al. 2022; An et al. 2018, 2018a). Because we felt that a dyad with a bidirectional relationship was more likely to depict a true referral relationship than one with a directed relationship (e.g., because two patient visit sequences would need to occur for reasons other than each physician making a deliberate choices to refer their patients to the other), we restricted the connections between physicians to mutual edges (those physician dyads with edges in both directions). Thus, the rule for an (undirected) edge to be present in the network is that at least one patient encounters one physician first and then the other and vice-versa for at least one other patient during 2014.

A network of 35,765 physicians resulted. However, following linkage of the physicians with the prescription drug data, the network was reduced in size as not all physicians prescribed drugs in at least one of the three risky drug classes (section "Final network and prescription drug analytic datasets"). The network methodology including ERGMs, measures of homophily and the identifiability of their effects, and the extra-dyadic network statistics and associated statistical tests are described in section "Network Modeling and Statistical Analysis Methods".

Overview of prescription drug data and methods

A distinct nationwide 40% random sample of Medicare Part D claims (prescription drug events) from 2014 was used to obtain beneficiaries' prescription fill records and identify their corresponding prescribers for opioids, benzodiazepines and sedative-hypnotics. The beneficiaries' prescription records, including the physicians who prescribed their drugs from Medicare Part D claims, were used to trace the trajectories of patients' prescriptions and then to construct the physicians' prescribing indexes. The construction of the physician prescribing indexes involves specialized algorithms for attributing the physician(s) responsible for both prescribing and deprescribing transitions that are presented and described in the Supplemental Appendix. The formation of drug-state transition matrices specific to each physician and the summarization of these matrices in the form of prescribing indexes to test for the presence of homophily in risky prescribing are described in section "Measures of homophily in physician prescribing and deprescribing".

Final network and prescription drug analytic datasets

As noted in section "Overview of physician network formation and methods", the 35,765 physicians in the Ohio shared-patient network were merged with the above dataset of physicians who prescribed at least one drug in the drug classes of interest (opioids, benzodiazepines, or sedative-hypnotics) in 2014 according to Medicare Part D data. This reduced the number of physicians in both the network and the prescription drug data sets to 22,655 physicians. Finally, we reduced the network to its largest connected component (LCC) to eliminate isolated dyads (pairs of physicians who only shared patients with each other and thus have a network degree of 1) as such physicians were likely practicing in a part-time or other reduced manner, resulting in network and prescription drug datasets containing 17,363 linked physicians (see Figure 1 in the Supplemental Appendix).

Network modeling and statistical analysis methods

Exponential random graph models (ERGMs)

An ERGM is an exponential family model in which the dependent variable is a sociocentric network of relationships and the predictors are statistics reflecting network features believed to underlie the network. Homophily is one such network feature making ERGMs a natural choice for modeling the association between the similarity of physician prescribing indexes and the likelihood that they are connected in the network. Standard regression models cannot handle sociocentric network data if the status of the edges (ties) in the network are statistically dependent as this violates the independence and

no-interference assumptions of standard regression (Contractor et al. 2006). ERGMs overcome this issue and allow nodal attributes, edge attributes, dyadic dependencies, and some higher-order network dependencies to be simultaneously modelled (Robins et al. 2007, 2007; Snijders et al. 2006). ERGMs model the probability distribution of all possible networks given a set of nodes, a discrete-valued distribution of a categorical random variable with a large number of possible states (O'Malley and Marsden 2008), and in estimation seek the parameters weighing the importance of each network statistic that make the observed network the most likely among all of its possible realizations (Robins et al. 2007). Mathematically, ERGMs model the probability that a random network $\mathbf{A}$ is realized by an observed network $\mathbf{a}$ ($a_{ij} = 1$ if there is an edge between nodes i and j and 0 otherwise) as:

$$\Pr(\mathbf{A} = \mathbf{a} \mid \mathbf{X} = \mathbf{x}) = \frac{1}{\kappa(\mathbf{x})} \exp\left\{\sum_{\mathrm{p}} \eta_{\mathrm{p}} \mathrm{g}_{\mathrm{p}}(\mathbf{a}, \mathbf{x})\right\} \tag{1}$$

where $\kappa(\mathbf{x}) = \sum_{\mathbf{a}\in\mathcal{A}} \exp\{\sum_p \eta_p g_p(\mathbf{a}, \mathbf{x})\}$ is a normalizing constant that makes the probabilities sum to 1 across possible networks with the given number of nodes and having attributes $\mathbf{X}$ with realized value $\mathbf{x}$ (Goodreau 2007; O'Malley and Marsden 2008); the vector of attributes of node i is the ith row of $\mathbf{x}$, denoted $\mathbf{x}_i^T$; and $g_p(\mathbf{a}, \mathbf{x})$ is the pth network statistic. In general, a positive η_p for the network configuration represented by $g_p(\mathbf{a}, \mathbf{x})$ indicates that the model for the network favors networks with feature p while a negative value indicates that networks with a high prevalence of feature p have lower likelihood.

Conditional interpretation of ERGM coefficients

The model in (1) implies that the conditional probability of a single edge satisfies:

$$\mathrm{logit}(\Pr(\mathrm{A_{ij}} = 1 \mid \mathbf{A}_{\mathrm{ij}}^{\mathrm{c}} = \mathbf{a}_{\mathrm{ij}}^{\mathrm{c}}, \mathbf{X} = \mathbf{x}) = \sum_{\mathrm{p}} \eta_{\mathrm{p}} \delta_{\mathrm{p}}(\mathbf{a}_{\mathrm{ij}}^{\mathrm{c}}, \mathbf{x}), \tag{2}$$

where $\mathbf{A}_{ij}^c$ denotes all edges in the network other than A_{ij} and

$$\delta_p(\mathbf{a}_{ij}^c, \mathbf{x}) = g_p(\mathbf{a}_{ij}^c \cup \{A_{ij} = 1\}, \mathbf{x}) - g_p(\mathbf{a}_{ij}^c \cup \{A_{ij} = 0\}, \mathbf{x})$$

is the difference in network statistic $g_p(\mathbf{a}, \mathbf{x})$ when $A_{ij} = 1$ to when $A_{ij} = 0$Hunter et al. (2008). Therefore, the parameter η_p is the log-odds of $A_{ij} = 1$ to $A_{ij} = 0$ if the presence of A_{ij} would lead to a one-unit increase in $g_p(\mathbf{a})$, conditioned on the rest of the network. In general, changing the value of A_{ij} from 0 to 1 can impact multiple network statistics, complicating the interpretation of the parameter values (see end of Sect. 3.3).

In theory, any network statistic may be included as a predictor in (1) and in practice a wide range of statistics capturing various network features have been represented in ERGMs (Morris et al. 2008; O'Malley and Marsden 2008). Commonly used network statistics include the number of: edges, reciprocated or mutual edges (for directed networks), certain degree-related configurations (e.g., k-stars), triadic configurations (e.g., triangles, transitive triads, three-cycles), node attributes (e.g., node factor and node covariates) and network statistics that quantify the level of homophily of specified attributes in the network. An example of the latter is the nodematch term, which

Table 1 Definitions of ERGM terms for undirected networks and interpretations

Terms	Math definition	Interpretation
Edges	$m = \frac{1}{2}\sum_{ij} a_{ij}$	Number of edges in the network; controls for network density
Node attribute terms		
nodefactor	$Z_k = \sum_{i<j}(I(x_i = k) + I(x_j = k))a_{ij}$	Number of times a node possessing a categorical attribute of value k appears on an edge in the network
nodecov	$Z = \sum_{i<j}(x_i + x_j)a_{ij}$	For continuous attributes, the sum of the attribute across node pairs for all edges present in the network
Homophily terms		
nodematch	$S = \sum_{i<j} I(x_i = x_j)a_{ij}$	Uniform homophily; the number of edges in which the two nodes have the same categorical attribute
nodematch	$S_l = \sum_{i<j} I(x_i = l)I(x_j = l)a_{ij}$	Differential homophily; the number of edges whose two nodes have the same categorical attribute value of l
absdiff	$S = \sum_{i<j} \lvert x_i - x_j \rvert a_{ij}$	For continuous attributes, the sum of absolute differences in the attribute within a dyad across all edges present in the network

$\mathbf{A} = [a_{ij}]$ is the adjacency matrix of the binary-undirected network and $a_{ij} = 1$ if physician i and j shared patients during 2014. The variables x_i and x_j are the node attributes of physician i and j, and k and l denote the values of a categorical attribute held by the two actors comprising a dyad

in its simplest form (see the uniform nodematch specification in Table 1) counts the number of edges in the network with identical values of a categorical attribute x, $nodematch_x(\mathbf{a}, \mathbf{x}) = \sum_{i<j} I(x_i = x_j)a_{ij}$ and so

$$\delta_m(\mathbf{a}_{ij}^c, \mathbf{x}) = I(x_i = x_j) \tag{3}$$

where m denotes *nodematch*. Therefore, if η_m denotes the coefficient of $nodematch_x(\mathbf{a}, \mathbf{x})$, η_m is the increase in the log-odds that $A_{ij} = 1$ to $A_{ij} = 0$ if $x_i = x_j$. Intuitively one can see that homophily in x is positive in the network if the probability of $A_{ij} = 1$ is greater when physicians i and j have the same x than if they don't.

The ergm package within in the statnet suite of packages in R contains an extensive list of network statistics (Hunter et al. 2008) and (as in the risky prescribing application) may be used to estimate ERGMs (Hunter et al. 2008; Handcock et al. 2008) while Table 1 includes the network statistics used in this paper.

Network statistics measuring homophily and the identifiability of their effects

Homophily can be thought of as a within-dyad interaction between the two nodes comprising the dyad of the given attribute (Hunter et al. 2008; Morris et al. 2008). When studying the homophily of an attribute, it is important to adjust for the main effect of the node attribute to ensure that homophily is a relative measure as opposed to being confounded with the prevalence of the attribute across the network. The network statistics associated with these main effects are named *nodefactor* (for categorical attributes) and *nodecov* (for continuous attributes) in ERGMs. Table 1 shows the mathematical definitions of the ERGM terms used in this study and their interpretations. For illustration, suppose that x_i is a scalar (i.e., each node in the network has a single attribute) and that the network is undirected. For a binary

node attribute x taking a value of 0 or 1, the network statistic for the *nodefactor* term is given by,

$$\sum_{i<j}(x_i + x_j)a_{ij}. \tag{4}$$

The two statistics added by the differential homophily *nodematch* term are,

$$\sum_{i<j} x_i x_j a_{ij} \tag{5}$$

for the node attribute taking a value of 1, and

$$\sum_{i<j}(1 - x_i)(1 - x_j)a_{ij} = \sum_{i<j}(1 - (x_i + x_j) + x_i x_j)a_{ij} \tag{6}$$

for the node attribute taking a value of 0 (Table 1). For a binary node attribute, the three predictors in Eqs. (4), (5) and (6) are linearly dependent and an ERGM for an undirected network including the *edges* term $\sum_{i<j} a_{ij}$, the *nodefactor* term and the two *nodematch* terms corresponding to differential network homophily is not identifiable. The lack of identifiability is seen by the fact that the *nodematch* term (Eq. 6) is the sum of the *edges* (Table 1), *nodefactor* (Eq. 4) and the 1-level *nodematch* (Eq. 5) terms. Hence, when controlling for network density with the *edges* term and the main effect of an attribute with a *nodefactor* term, the uniform homophily statistic $\sum_{i<j} I(x_i = x_j) = \sum_{i<j}((1 - x_i)(1 - x_j) + x_i x_j)a_{ij}$ (the sum of the two *nodematch* terms) can be identified but both of its components (the two differential homophily terms) cannot. When estimating homophily while adjusting for the node-level effect of the same attributes, only the effect of uniform homophily (the effect of having the same value of the attribute irrespective of its value) can be estimated. In contrast, under differential homophily, the coefficients of the network statistics are unidentifiable due to linear dependencies between the predictors.

Representation of homophily as an interaction effect

When an ERGM includes the density, nodefactor(x) and uniform nodematch(x) terms with associated coefficients η_d, η_f and η_m, respectively, and x is a binary attribute, it follows that

$$\mathrm{logit}(\Pr(\mathrm{A_{ij}} = 1 \mid \mathbf{A}^{\mathrm{c}}_{\mathrm{ij}} = \mathbf{a}^{\mathrm{c}}_{\mathrm{ij}}, \mathbf{X} = \mathbf{x}) = \eta_{\mathrm{d}} + \eta_{\mathrm{f}}(\mathrm{x_i} + \mathrm{x_j}) + \eta_{\mathrm{m}}\mathrm{I}(\mathrm{x_i} = \mathrm{x_j}), \tag{7}$$

as $\delta_d(\mathbf{a}^c_{ij}, \mathbf{x}) = 1$, $\delta_f(\mathbf{a}^c_{ij}, \mathbf{x}) = x_i + x_j$, and $\delta_m(\mathbf{a}^c_{ij}, \mathbf{x}) = I(x_i = x_j)$. Forming a system of four equations by evaluating (7) at the four combinations of (x_i, x_j) and solving for the homophily coefficient η_m, it follows that:

$$\exp(\eta_m) = \frac{\Pr(\mathrm{A_{ij}} = 1 \mid \mathrm{x_i} = 1, \mathrm{x_j} = 1, \mathbf{A}^{\mathrm{c}}_{\mathrm{ij}} = \mathbf{a}^{\mathrm{c}}_{\mathrm{ij}})\Pr(\mathrm{A_{ij}} = 1 \mid \mathrm{x_i} = 0, \mathrm{x_j} = 0, \mathbf{A}^{\mathrm{c}}_{\mathrm{ij}} = \mathbf{a}^{\mathrm{c}}_{\mathrm{ij}})}{\Pr(\mathrm{A_{ij}} = 1 \mid \mathrm{x_i} = 0, \mathrm{x_j} = 1, \mathbf{A}^{\mathrm{c}}_{\mathrm{ij}} = \mathbf{a}^{\mathrm{c}}_{\mathrm{ij}})\Pr(\mathrm{A_{ij}} = 1 \mid \mathrm{x_i} = 1, \mathrm{x_j} = 0, \mathbf{A}^{\mathrm{c}}_{\mathrm{ij}} = \mathbf{a}^{\mathrm{c}}_{\mathrm{ij}})}. \tag{8}$$

Recognizing the right-hand-side of (8) as a contrast of a multiplicative interaction, η_m is seen to be an interaction term capturing the incremental effect of a pair of actors sharing the same value of the node attribute x above and beyond the density of the edges and the distribution of x across the network. The same interpretation holds if the ERGM includes other nodefactor, nodecov, nodematch and absdiff terms and a similar interpretation holds for nodecov terms; all homophily terms have interpretations emulating those of interaction effects.

Triadic homophily network statistics

ERGMs including only the network statistics discussed to date are examples of dyadic independent models. These can be estimated straightforwardly. However, because the status of one of the three triads comprising a tetrad restricts the possible status of the triad with which it shares an edge, network statistics for triadic terms restrict the parameter-space of an ERGM—at a minimum inducing statistical dependencies in estimation and at worst leading to degeneracy. To avoid these issues, we computed two triadic statistics that are restricted through the involvement of attribute information:

- The proportion of closed triangles with the same node attribute, $Tri_1(\mathbf{a}, \mathbf{x})$.
- The proportion of open two-paths (2-stars or open-triangles) with the same node attribute that are closed, $Tri_2(\mathbf{a}, \mathbf{x})$.

$Tri_1(\mathbf{a}, \mathbf{x})$ is the proportion of closed triangles in which each of the three actors has the attribute in common while $Tri_2(\mathbf{a}, \mathbf{x})$ is the proportion of two-paths (2-stars or open triangles) with the same node attribute that are closed, an attribute-specific version of transitivity (Latapy et al. 2008). For a binary node attribute (taking the value of 0 or 1), $Tri_1(\mathbf{a}, \mathbf{x})$ is defined as,

$$Tri_1(\mathbf{a}, \mathbf{x}) = \frac{\Sigma x_i x_j x_k \cdot a_{ij} a_{jk} a_{ki}}{\Sigma a_{ij} a_{jk} a_{ki}}. \tag{9}$$

Likewise, for a binary node attribute, the statistic Tri_2 is defined as,

$$Tri_2(\mathbf{a}, \mathbf{x}) = \frac{\Sigma x_i x_j x_k \cdot a_{ij} a_{jk} a_{ki}}{\Sigma x_i x_j x_k \cdot a_{ij} a_{ik}}. \tag{10}$$

In our application, $\mathbf{A} = [a_{ij}]$ is the adjacency matrix of the binary-undirected network (see Sect. 2.1) such that $a_{ij} = 1$ if physician i and j shared at least one patient in each direction (i visited before j and vice-versa) during 2014. Thus, $Tri_1(\mathbf{a}, \mathbf{x})$ is the proportion of times that three physicians who shared patients among themselves all contributed to risky prescribing, while $Tri_2(\mathbf{a}, \mathbf{x})$ is the proportion of 2-star configurations with physician i as the apex (triads with an undirected 2-path from j to k via i) that are closed (physicians j and k also shared patients) among those for which nodes i, j, and k are all risky prescribers. Therefore, $Tri_2(\mathbf{a}, \mathbf{x})$ is an attribute-restricted version of node transitivity (Latapy et al. 2008). See Fig. 1 for an illustrative example of $Tri_1(\mathbf{a}, \mathbf{x})$ and $Tri_2(\mathbf{a}, \mathbf{x})$.

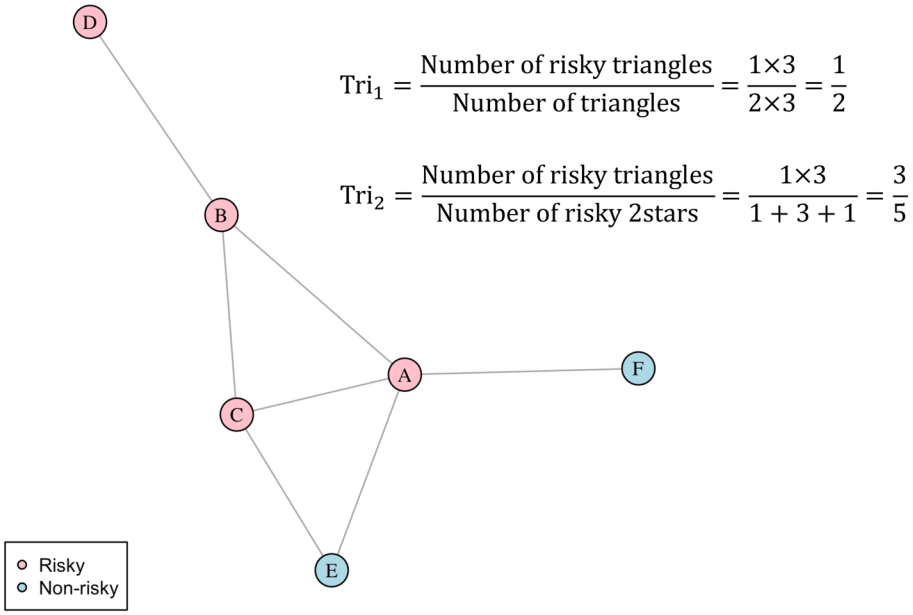


Fig. 1 Illustrative computation of triadic homophily statistics $Tri_1(\mathbf{a}, \mathbf{x})$ and $Tri_2(\mathbf{a}, \mathbf{x})$. Suppose nodes A, B, C, and D are physicians who have contributed to risky prescribing, and nodes E and F are non-risky-prescribing physicians. The number of risky 2-stars with nodes A, B, C, and D being the center vertex is 1, 3, 1, and 0, respectively. Therefore, the total number of 2-stars among risky prescribing physicians is five

Non-parametric test for triadic homophily

The numerator and denominator in (9) have their respective ERGM terms in the ergm package (Handcock et al. 2008). However, the ratio of them is not available in ergm. Likewise, the denominator in (10), the total number of 2-stars among nodes with a certain attribute, is not directly available in ergm. Instead of estimating an ERGM, we perform non-parametric tests by randomly re-distributing the node attribute in question across the nodes, preserving the total number of nodes, the number of nodes with a certain attribute, and the structure of the observed network. For example, in our application, we performed 30 random permutations of the attribute of interest across the nodes of the network and computed $Tri_1(\mathbf{a}, \mathbf{x})$ and $Tri_2(\mathbf{a}, \mathbf{x})$ on each. The resulting empirical distribution of these test statistics is evaluated on each of the 30 permuted networks to evaluate the plausibility of the observed value of the statistic under the null hypothesis of no homophily. Tri_1 and Tri_2 generalize to continuous node attributes by standardizing the attribute to have a range from 0 and 1. All the analyses were performed using Python 3.7 and R (R Core Team 2022; Van Rossum and Drake 2009).

Measures of homophily in physician prescribing and deprescribing

In this section, we describe the construction of measures of physician prescribing and deprescribing that serve as the focal node attributes throughout the empirical analysis of the motivating risky prescribing application. In section "Results of network-related risky prescribing analysis", the resulting measures will be analyzed descriptively, included in ERGMs as nodematch or absolute difference (absdiff) terms (Table 1) to study the homophily of prescribing accounting for the density of network connections and the levels of risky prescribing across the network (via the nodefactor or nodecov terms in

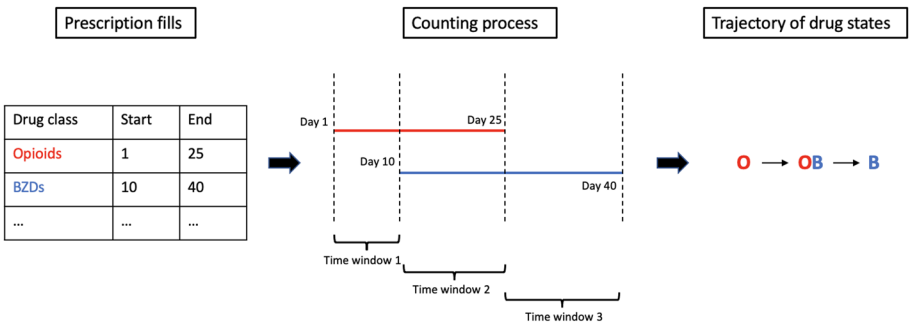


Fig. 2 Workflow of representing patient prescription states. Note: The left-hand panel (L) shows a made-up example of a patient's sequence of prescription fills with their corresponding drug class. The center panel (C) shows the counting process to split the sequence of prescription fills into discrete exposure time intervals that reflect the initialization and the discontinuation of a prescription fill. The red line indicates the prescription fill length of the opioid in panel (L), and the blue line indicates the benzodiazepine (BZD) fill length. The right-hand panel (R) shows the corresponding prescription state during each time interval in panel (C) and the transition between them, forming a trajectory of prescription states across time. "O" stands for filling an opioid, "B" stands for filling a BZD, and "OB" stands for filling an opioid and a BZD concurrently

Table 1), and may be used as the attribute measure in the tests of extra-dyadic homophily developed in section "Triadic homophily network statistics".

Modeling patient prescription states

For each patient, their prescription fills of the three-drug classes of interest were divided into discrete time intervals, capturing the initialization and discontinuation of a prescription, with each time interval reflecting the number and class of drugs they were prescribed (Fig. 2). Because every initialization or discontinuation of a drug changes a patient's prescription state, it is important to distinguish new fills from refills. As detailed in Section 2.1 of the Supplemental Appendix, we implemented an empirical rule of 20% overlapping fill length, where a subsequent prescription fill of the same drug was appended to the preceding fill if the gap in between was less than 20% of the fill length of the preceding prescription.

Following the determination of discrete patient prescription state time-intervals for each patient, we assigned each interval to one of the $2^3 = 8$ combinations of prescription states in the set {zero, O, B, S, OB, OS, BS, OBS}, where O, B, and S denote opioids, benzodiazepines and sedative hypnotics, respectively. State *zero* is a state of taking no drugs in the three-drug classes of interest. States $\{O, B, S\}$ correspond to taking at least one drug in precisely one of the three drug classes. States $\{OB, OS, BS\}$ correspond to taking drugs in at least two different drug classes concurrently. State *OBS* indicates concurrent receipt of at least one drug in each of the three different classes.

For the ease of mathematical notation, the eight prescription states are numbered 1 to 8. Whereas we consider a patient taking drugs in multiple drug-groups as in a riskier state than a patient taking only one, the drug-groups are not themselves ordered based on severity. Herein we consider a lone benzodiazepine to have the same risk as a lone opioid. However, any combination of a benzodiazepine and an opioid is considered to have the same risk as the combination of a sedative-hypnotic and an opioid but any combination of two drug states (states 5, 6 and 7) is considered more risky than any of the three singleton drug states (2, 3, and 4).

Attributing physician responsibility to prescribing and deprescribing prescriptions

Enumerating the transitions between the eight patient drug states {zero, O, B, S, OB, OS, BS, OBS} over all patients yields an 8 by 8 transition matrix whose off-diagonal elements contain the number of transitions from one state to another over the entire sample. However, to form physician-specific prescribing indexes, physician-specific versions of this matrix were required, necessitating the attribution of each patient drug state transition to one or more *responsible* physician(s). We describe the attribution processes when a patient starts or stops taking drugs in one of the $\{O, B, S\}$ drug-categories in Sections 2.2 and 2.3 of the Supplemental Appendix, respectively. While a physician ID appears with each prescription in Medicare Part D, no such entry is made for deprescribing making the identification and attribution of deprescribing events particularly intricate, as evinced by the two pseudo-codes constructed for this purpose (see Algorithms 1 and 2 in the Supplemental Appendix).

Physician transition responsibility matrix

Following the attribution of each patient drug state transition to one or more responsible physicians, we constructed an 8 by 8 physician transition responsibility count matrix (*PTRCM*) for each physician using all of the transitions in patient drug status attributed to them. The rows of the *PTRCM* correspond to patient prescription states in the preceding time interval, while the columns correspond to patients' prescription states in the current time intervals. For physician k, the PTRCM is given by

$$PTRCM^{(k)} = \begin{bmatrix} C_{1,1}^{(k)} & C_{1,2}^{(k)} & \dots & C_{1,j}^{(k)} & \dots & C_{1,8}^{(k)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ C_{i,1}^{(k)} & C_{i,2}^{(k)} & \dots & C_{i,j}^{(k)} & \dots & C_{i,8}^{(k)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ C_{8,1}^{(k)} & C_{8,2}^{(k)} & \dots & C_{8,j}^{(k)} & \dots & C_{8,8}^{(k)} \end{bmatrix}. \tag{11}$$

where $C_{i,j}^{(k)}$ is the total number of patient prescription state transitions from (prescription) state i to state j for which physician k was deemed responsible. To mathematically depict the calculation of $C_{i,j}^{(k)}$, let h denote a patient, s the prescription change occasion, D_{hs} patient h's prescription state after prescription change s, and P_{hs} the corresponding responsible physician. Then

$$C_{i,j}^{(k)} = \sum_h \sum_s \mathbf{I}(D_{hs} = i, D_{h(s+1)} = j)\mathbf{I}(P_{hs} = k). \tag{12}$$

where $\mathbf{I}(event) = 1$ if *event* is true and 0 otherwise. To account for the scenario when multiple physicians are responsible for a prescription state transition from i to j, let N_{hs} denote the total number of responsible physicians for prescription state transition s of patient h and define

$$C_{i,j}^{(k)} = \sum_h \sum_s \mathbf{I}(D_{hs} = i, D_{h(s+1)} = j)\frac{\sum_r \mathbf{I}(P_{hsr} = k)}{N_{hs}}. \tag{13}$$

Quantitative measures of physician prescribing behavior

Motivated by aspects of decision-making during physicians' prescribing practice, in this section we define multiple summary measures of $C_{i,j}^{(k)}$ termed prescribing indexes. As noted at the start of section "Measures of homophily in physician prescribing and deprescribing", these prescribing indexes are the physician attributes we use to test for homophily using ERGMs and to test for extra-dyadic homophily.

Difference in relative frequency of prescribing and deprescribing

As a first prescribing index, we construct a measure of the relative net tendency of a provider to prescribe new classes of drugs over their tendency to remove classes of drugs. Let $\mathbf{I}_{num}(\omega)$ denote the number of drugs a patient is taking while in prescription state ω, so that

$$\mathbf{I}_{num}(\omega) = \begin{cases} 0, & \omega = 1, \\ 1, & \omega \in \{2,3,4\}, \\ 2, & \omega \in \{5,6,7\}, \\ 3, & \omega = 8. \end{cases} \tag{14}$$

Then, for example, a transition from state *B* (or state 3) to state *OBS* (or state 8) involves signed and absolute changes of 2 drugs whereas the transition in the opposite direction involves a signed change of -2 drugs and an absolute change of 2 drugs. We then define the overall net prescribing-deprescribing index for physician k as,

$$I_1^{(k)} = \frac{\sum_j \sum_{i \neq j} C_{i,j}^{(k)} (\mathbf{I}_{num}(i) - \mathbf{I}_{num}(j))}{\sum_j \sum_{i \neq j} C_{i,j}^{(k)} |\mathbf{I}_{num}(i) - \mathbf{I}_{num}(j)|}, \tag{15}$$

a quantity that ranges between -1 and 1. An integral element of $I_1^{(k)}$ is that the counts $C_{i,j}^{(k)}$ of the number of transitions to which physician k contributed are multiplied by the number of drug changes involved. A more general measure is obtained by introducing a parameter α that weighs the number of drug changes, as follows:

$$I_\alpha^{(k)} = \frac{\sum_j \sum_{i<j} C_{i,j}^{(k)} |\mathbf{I}_{num}(i) - \mathbf{I}_{num}(j)|^\alpha - \sum_i \sum_{i>j} C_{i,j}^{(k)} |\mathbf{I}_{num}(i) - \mathbf{I}_{num}(j)|^\alpha}{\sum_j \sum_{i \neq j} C_{i,j}^{(k)} |\mathbf{I}_{num}(i) - \mathbf{I}_{num}(j)|^\alpha}, \tag{16}$$

for $\alpha \geq 0$. If $\alpha = 1$ we obtain (15) while under the boundary case $\alpha = 0$ yields the reduced measure, $I_0^{(k)}$, in which the numerator is the sum of the above-diagonal elements of $C^{(k)}$ less the sum of the below-diagonal elements.

Frequency of transitioning to riskiest prescribing state

We next focus on physician k's tendency to write prescriptions that transition patients to the riskiest form of prescribing, state *OBS* (or state 8), as this may best capture their willingness to expose patients to risk. This measure is given mathematically as,

$$I_{OBS}^{(k)} = \frac{\sum_{i<8} C_{i,8}^{(k)}}{\sum_j \sum_{i \neq j} C_{i,j}^{(k)}} \tag{17}$$

where $C_{i,8}^{(k)}$ is the number of transitions physician k contributed to for which patients transitioned from state i to state 8 (state *OBS*). We construct a reduced version of $I_{OBS}^{(k)}$

by thresholding the numerator of (17) at 0 to obtain the binary indicator of whether physician k ever contributed to bringing a patient into state *OBS*:

$$I^{(k)}_{everOBS} = \mathbf{I}\left(\sum_{i<8} C^{(k)}_{i,8} > 0\right). \tag{18}$$

Frequency of substantial prescribing and deprescribing transitions

We now develop measures that quantify the percentage of patients' prescription drug state transitions physician k contributed to involving two or more drug class additions or removals. For example, a transition from state *zero* to state *OBS* involves three additions, and a transition from state *OBS* to state *B* involves two removals of drug classes, respectively. The intent of these measures is to focus on the frequency with which a physician makes substantial changes to their patients' drug regimes. We construct separate prescribing and deprescribing measures as these may capture different aspects of a physicians practice. The prescribing version of this measure is operationalized as the proportion of physician k's prescribing that involves the addition of two or more of the three targeted drug classes, given by

$$I^{(k)}_{presc2mr} = \frac{\sum_j \sum_{i<j} C^{(k)}_{i,j} \mathbf{I}(|\mathbf{I}_{num}(i) - \mathbf{I}_{num}(j)| \geq 2)}{\sum_j \sum_{i<j} C^{(k)}_{i,j}}, \tag{19}$$

where $\sum_j \sum_{i<j} C^{(k)}_{i,j} > 0$ as by construction of the study sample $\sum_j C^{(k)}_{0,j} > 0$. Similarly, the proportion of physician k's deprescribing that involves removing two or more drug classes is,

$$I^{(k)}_{deprescr2mr} = \frac{\sum_i \sum_{i>j} C^{(k)}_{i,j} \mathbf{I}(|\mathbf{I}_{num}(i) - \mathbf{I}_{num}(j)| \geq 2)}{\sum_i \sum_{i>j} C^{(k)}_{i,j}}, \tag{20}$$

if $\sum_i \sum_{i>j} C^{(k)}_{i,j} > 0$ and 0 otherwise.

Results of network-related risky prescribing analysis

We first present descriptive features of the physician network for the state of Ohio and the prescribing indexes used to study the presence of homophily and extra-dyadic homophily in this network in section "Physician shared-patient networks" and "Prescribing and deprescribing measures", respectively. The results for the ERGMs used to test for homophily while adjusting for the prevalence of edges in the network and the physicians' overall prescribing tendencies are in section "ERGMs for estimating independent effects of homophily" while the results of the tests of extra-dyadic homophily are in section "Triadic-level hyper homophily".

Physician shared-patient networks

Table 2 shows the network statistics of the entire (Ohio) shared-patient physician network, the prescribing network, and the LCC of the prescribing network. Around 63%

Table 2 Network statistics of the largest connected component of the shared-patient physician prescribing network (specific to prescriptions of opioids, benzodiazepines, and sedative-hypnotics) for Ohio in 2014

Descriptive statistics	Whole network	Prescribing network	LCC of prescribing network
Shared-patient physician network			
Number of nodes	35765	22655	17363
Number of ties	494462	265112	261816
Density	0.0008	0.0010	0.0010
Number of components	3002	2056	1
Size of LCC	27503	17363	17363
Degree (mean, IQR, SD)	27.7 (n/a, 44.0) (37.5)	23.4 (n/a, 38.0) (28.7)	30.2 (n/a, 45.0) (29.6)
Global clustering	0.168	0.171	0.171
Average path length	4.663	4.599	4.599
Volume (mean, IQR, SD)	62.7 (n/a, 96.0) (86.3)	70.6 (n/a, 112.0) (87.8)	91.6 (18.0, 136.0) (90.3)
Physician prescribing measures			
I_0 (mean, IQR)		0.871 (0.843, 1.0)	0.876 (0.875, 1.0)
I_{OBS} (mean, IQR)		0.009 (0.0, 0.0)	0.009 (0.0, 0.0)
$I_{everOBS}$ (# of 1, # of 0)		(1972, 20683)	(1412, 15951)
$I_{presc2mr}$ (mean, IQR)		0.030 (0.0, 0.0)	0.029 (0.0, 0.0)
$I_{deprescr2mr}$ (mean, IQR)		0.017 (0.0, 0.0)	0.016 (0.0, 0.0)

The physician network is constructed based on the overlap of patient care at any point during 2014 between physician dyads treating patients residing in Ohio. The prescribing network is a subset of the whole network where its physicians have prescribed at least one opioid, benzodiazepine, or sedative-hypnotic during 2014. Volume is the number of Ohio Medicare fee-for-service beneficiaries a physician encountered throughout 2014. The entries of n/a signify that the true value is suppressed to satisfy data suppression rules designed to protect patient privacy by the Center for Medicare and Medicaid Services. LCC = largest connected component

(22,655 out of 35,765) of physicians in the shared-patient physician network were identified as prescribers of at least one opioid, benzodiazepine, or sedative-hypnotic, and around half of the ties in the network took place among the prescribers. The LCC of the prescribing network consists of more than 76% of physicians and more than 98% of the ties in the full prescribing network. As shown in the final two rows of the upper segment of Table 2, the prescribing network and its LCC were similar in terms of network statistics and physician prescribing measures, except that physicians in the LCC had a slightly higher average node degree (just above 30) and the number of Ohio patients encountered annually (nearly 92 patients on average). The facts that the average degree and average volume of patients are both higher within the LCC of the network is reflective of the fact that the LCC retains on average the more highly connected physicians who see more patients.

Prescribing and deprescribing measures

We first present the network statistics of the entire Ohio shared-patient physician network, the prescribing network and the LCC of the prescribing network. In the LCC of the Ohio physician prescribing network, the distributions of the I_0 index of the relative net difference between prescribing and deprescribing, is skewed to the left often obtaining the largest possible value of 1 limiting the utility of this measure (Table 2). On average, among all the patient prescription state transitions a physician contributed to, around 0.9% of them involve bringing patients to state *OBS*, the riskiest state. Around

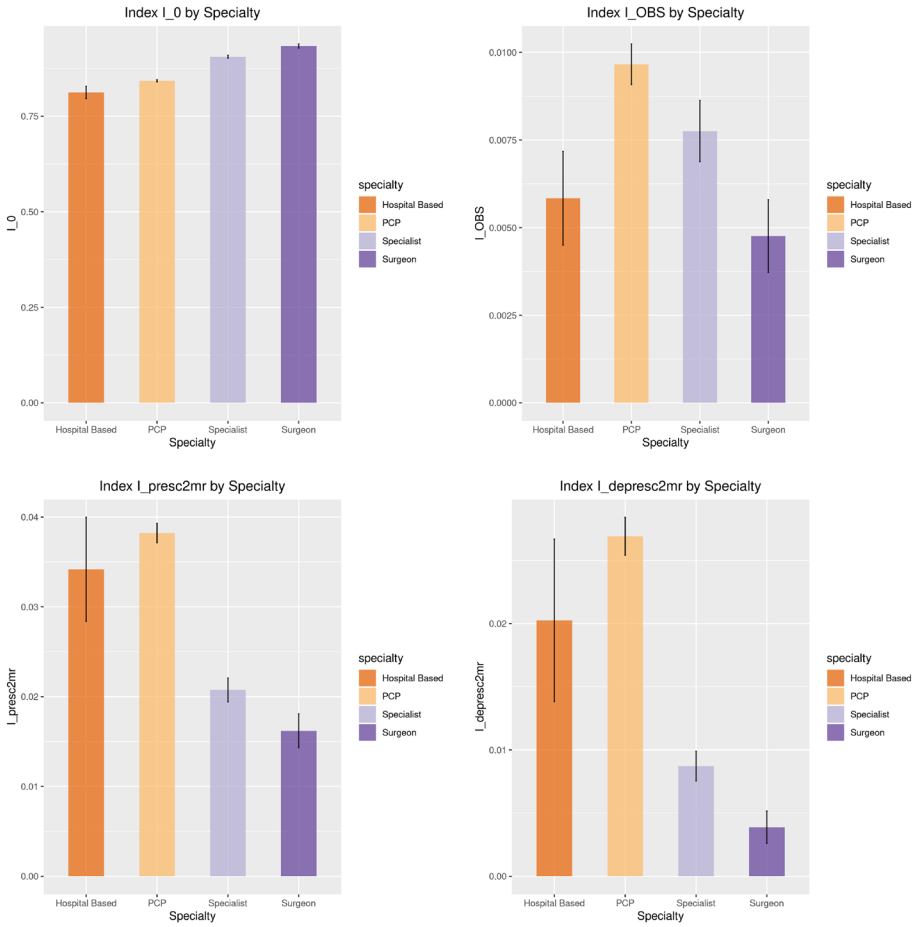


Fig. 3 Prescribing measures by specialty of physicians in the largest connected component of the shared-patient prescribing physician Ohio network in 2014. Specialists are medical specialists other than surgeons. Hospital-based services include anesthesiology, radiology, and pathology. PCP denotes primary care physicians. l_0 is the net relative difference in prescribing and deprescribing transitions (l_1, a measure in the same family as l_0 that accounts for the change in the number of prescription classes (Eq. 15), had a similar barplot), l_{OBS} is the prescribing index based on a physician's contribution to bringing patients to the riskiest prescription state *OBS*, $l_{presc2mr}$ ($l_{deprescr2mr}$) is the prescribing index based on a physician's contribution to prescribing (deprescribing) two or more drug types to patients. Error bars show the standard errors of the respective measures

8.1% of physicians (1,412 out of the sum of 1,412 and 15,921) have at least once contributed to a patient's transition to state *OBS*. Among all of the transitions associated with prescribing, around 2.9% of them involved adding two or more drugs. Likewise, among all the transitions associated with deprescribing, on average around 1.6% involved a reduction of two or more drugs. These proportions are nearly invariant between the prescribing network and its LCC, suggesting that little is lost from using the easier to analyze LCC as the basis of our network analyses.

Figure 3 presents mean values of four prescribing indexes by physician specialty for the LCC of the Ohio shared-patient physician network in 2014; each physician was classed as either a primary care physician, medical specialist or surgeon specialist based on their lookup information in the National Plan and Provider Enumeration System (NPPES) (Centers for Medicare & Medicaid Services 2012). In terms of overall prescribing and

deprescribing reflected by I_0, there was minimal difference across specialties, although surgeons and medical specialists appeared to have slightly higher average I_0 values than other specialties. Other prescribing measures reflect physicians' prescribing behavior with more granularity. Hospital-based physicians (often referred to as hospitalists) and primary care physicians (PCPs), in particular, have a higher likelihood of bringing patients to state OBS, prescribing two or more drugs, and deprescribing two or more drugs, compared to medical specialists and surgeons. While the relative magnitude of these differences is large, the absolute magnitude is modest as the three types of transitions at the core of these measures are at best infrequent and in the case of I_{OBS} occur seldomly. The confidence interval for hospitalists is the widest, implying that such physicians are the least common.

Section 3 of the Supplemental Appendix presents a visualization of the physician network that provides visual evidence of these homophily patterns. An important observation is that central physicians with higher patient volume and higher node degrees have lower risky prescribing intensity than peripheral physicians. In addition, there are closely positioned clusters of physicians with similar prescribing intensity and behavior.

ERGMs for estimating independent effects of homophily

The key terms of interest in the ERGMs for relating the physician network to the prescribing indexes are the estimated coefficients of the prescribing indexes; these estimate the level of homophily of physician (risky) prescribing in the LCC of the Ohio physician network. In line with the identifiability conditions derived in Sect. 3.3, we only include uniform homophily terms in the models. However, we fit four separate models corresponding to the separate inclusion of $I_{everOBS}$, I_{OBS}, $I_{presc2mr}$ and $I_{deprescr2mr}$ in the model and in each model adjust for network density, the prevalence of each physician specialty (*nodefactor* term), uniform homophily of physician speciality (*nodematch* term), and the physician-level distribution of the prescribing index (depending on the measure, *nodefactor* or *nodecov*).

Table 3 shows the estimated ERGM-adjusted homophily effects in the LCC of the shared-patient prescribing physician network. When controlling for network density and the main effects of nodal prescribing and deprescribing attributes, the network exhibited assortative clustering in terms of different prescribing measures. An overall homophily effect was found among physicians in ever bringing patients to the OBS state ($est. = 0.037, p < 0.001$). As illustrated by Eqs. (2) and (3), 0.037 is the log-odds that $A_{ij} = 1$ to $A_{ij} = 0$ conditional on $I_{everOBS,i} = I_{everOBS,j}$ and $\mathbf{A}^c_{ij}$, the remainder of the network. In addition, $\exp(0.037) = 1.038$ is the value of the multiplicative interaction effect given in Eq. 7 implying that net of the density of edges and the prevalence of $I_{everOBS}$ in the network, the likelihood of the edge A_{ij} being present is approximately 3.8% greater (a substantial number of edges given the substantial size of the network) if $I_{everOBS,i} = I_{everOBS,j}$ than otherwise.

The status of connections among physicians were also associated with the continuous prescribing measures; because these are distance measures, a negative coefficient estimate implies greater homophily. Physicians with a larger difference in their likelihood of transitioning patients to OBS were less likely to be connected to each other

Table 3 ERGM adjusted homophily effects for models estimated on the largest connected component of the Ohio 2014 shared-patient physician prescribing network

	Model 1			Model 2			Model 3			Model 4		
Network statistic	**Est.**	**SE**	***p***	**Est**	**SE**	***p***	**Est**	**SE**	***p***	**Est**	**SE**	***p***
Edges	− 5.732	0.012	***	− 5.614	0.006	***	− 5.609	0.006	***	− 5.620	0.006	***
Node attributes: prescribing indexes												
Binary-valued index (nodefactor term)												
Ever OBS ($l_{everOBS}$)	0.327	0.010	***									
Continuously-valued index (nodecov term)												
Proportion OBS (l_{OBS})				1.202	0.109	***						
Presc. 2 classes ($l_{presc2mr}$)							0.492	0.049	***			
Depresc. 2 classes ($l_{depresc2mr}$)										0.337	0.066	***
Node attributes: specialty												
Primary care (reference)												
Emergency medicine	− 0.584	0.006	***	− 0.614	0.006	***	− 0.614	0.006	***	− 0.611	0.006	***
Neurologist	− 0.338	0.010	***	− 0.358	0.010	***	− 0.359	0.010	***	− 0.357	0.010	***
Psychiatrist	− 0.668	0.009	***	− 0.654	0.009	***	− 0.653	0.009	***	− 0.651	0.009	***
Other	− 0.396	0.004	***	− 0.423	0.005	***	− 0.423	0.005	***	− 0.420	0.005	***
Homophily: prescribing indexes												
Binary-valued index (nodematch term: 1 if physicians have same value of index and 0 otherwise)												
Ever OBS ($l_{everOBS}$)	0.037	0.011	***									
Continuously-valued index (absdiff term: the more similar the index the smaller the difference)												

Table 3 (continued)

Network statistic	Model 1			Model 2			Model 3			Model 4		
	Est.	SE	*p*	Est	SE	*p*	Est	SE	*p*	Est	SE	*p*
Proportion OBS (l_{OBS})				− 1.200	0.114	***						
Presc. 2 classes ($l_{presc2mr}$)							− 0.619	0.054	***			
Depresc. 2 classes ($l_{deprescr2mr}$)										− 0.203	0.068	**
Homophily: specialty												
Primary care	− 1.509	0.009	***	− 1.509	0.009	***	− 1.511	0.009	***	− 1.509	0.009	***
Emergency medicine	0.541	0.019	***	0.541	0.019	***	0.541	0.019	***	0.541	0.019	***
Neurologist	− 0.190	0.091	*	− 0.190	0.091	*	− 0.190	0.091	*	− 0.190	0.091	*
Psychiatrist	0.673	0.050	***	0.670	0.050	***	0.674	0.050	***	0.673	0.050	***

The node attribute term and homophily term for an attribute were in separate models for each of the prescribing indexes, yielding four models. In each model, physician specialty and homophily of physician specialty (restricted to uniform effects across the different specialties) were included in the model. Absdiff is the ergm package term for examining the homophily of a continuous node attribute, with a negative estimate indicating positive homophily (smaller differences imply greater likelihood of a network connection). I_{OBS} is the prescribing index based on a physician's contribution to bringing patients to the riskiest prescription state *OBS*, $I_{presc2mr}$ ($I_{deprescr2mr}$) is the prescribing index based on a physician's contribution to prescribing (deprescribing) two or more drug types to patients.

PC primary care

Significance levels: $***p < 0.001; **p < 0.01; *p < 0.05$

($est. = -1.200, p < 0.001$) while those with larger differences in the likelihood of prescribing two or more drugs to patients at once had a lower likelihood of a tie ($est. = -0.619, p < 0.001$). Similarly, ties were less likely if there was a larger difference in their propensity to deprescribe two or more drugs, although the effect is substantially lower in magnitude ($est. = -0.203, p < 0.01$). The last two results justify the consideration of deprescribing separately from prescribing. Because a physician's deprescribing behavior may address different aspects of their medical practice than their prescribing behavior and the act of prescribing is more formerly recognized in terms of generating a medical claim, it is reasonable to expect that the prescribing homophily coefficient would exceed that of the deprescribing homophily coefficient for the analogous measure. Observing that the estimated coefficient of $I_{presc2mr}$ is approximately 3 times that of $I_{deprescr2mr}$ is consistent with this hypothesis and so is a sanity check that our results cohere with intuitive reasoning and thus as a form of face validity of our approach.

The estimated effect of physicians' propensity to form ties with other physicians of the same specialty was consistent across all models. Compared to PCPs, emergency medicine physicians, neurologists, and psychiatrists had fewer connections with other physicians in the network. After controlling for the main effects of each type of physician specialty, PCPs and neurologists were less likely to be connected to same specialty physicians. In contrast, emergency medicine physicians and psychiatrists were more likely to form ties with physicians of the same specialty. We believe the latter finding reveals that psychiatrists are more likely to send a patient to another psychiatrist for a second opinion or that patients are more likely to doctor-shop among psychiatrists than among other specialists and especially PCPs. In contrast, PCPs seldom refer patients to other PCPs. Finally, the positive main-effects of the prescribing indices and the negative main-effects for each type of speciality reflects that physicians involved in more prescription transitions have more ties in the network and that PCPs have more network ties to other physicians than do specialists.

A goodness-of-fit analysis revealed that each of the models fit the data well with no indication of a major source of lack-of-fit (see Section 4 of the Supplemental Appendix). Furthermore, due to the inclusion of only dyadic independence terms as network statistics in each of the ERGMs, there are no concerns of model degeneracy and so we have near to 100% assurance that the algorithm for estimating each of the ERGMs converged.

When the ERGM analysis was stratified by health referral region (HRR), we found sizeable heterogeneity in the level of homophily (see Section 5 of the Supplemental Appendix for details and results of these analyses). However, the signs of the homophily terms were almost exclusively negative implying that the results of the ERGM analysis reported in Table 3 hold relatively generally.

Triadic-level hyper homophily

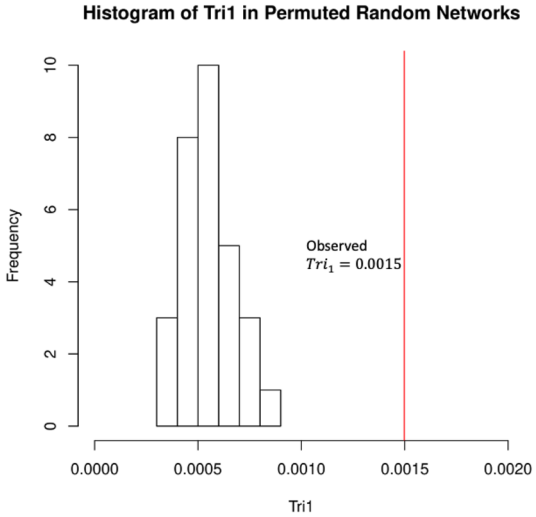
In this section, we report the results from the non-parametric permutation test of extra-dyadic homophily developed in section "Triadic homophily network statistics" and illustrated in Fig. 1. Because it obtained the estimated effect of greatest magnitude across the ERGMs in section "Exponential random graph models (ERGMs)", we performed the test for the $I_{everOBS}$ prescribing index. Hence, using the notation

of section "Triadic homophily network statistics", $\mathbf{x} = (I^{(1)}_{everOBS}, I^{(1)}_{everOBS}, \ldots, I^{(N)}_{everOBS})^T$, where $N = 17,363$ is the total number of physicians in the LCC of the Ohio network. The realized values of the triad-level measures of homophily quantified by the network statistics $Tri_1(\mathbf{a}, \mathbf{x})$ and $Tri_2(\mathbf{a}, \mathbf{x})$ for the $I_{everOBS}$ risky prescribing index are 0.0015 and 0.0007, respectively (Fig. 4). The interpretation of the realized $Tri_1(\mathbf{a}, \mathbf{x})$ is that among 10,000 closed triangles (mutual patient-sharing within all three physician dyads), 15 of them include nodes with the same attribute (the three physicians each contributed to bringing at least one patient to the riskiest prescription state *OBS*). The interpretation of $Tri_2(\mathbf{a}, \mathbf{x})$ is that among 10,000 open two-paths (2-stars) with the same node attribute ($I_{everOBS}$) in the network, 7 of them are closed. The attribution re-distribution permutation test described in section "Non-parametric test for triadic homophily" found that the observed values of $Tri_1(\mathbf{a}, \mathbf{x})$ and $Tri_2(\mathbf{a}, \mathbf{x})$ in the network were significantly higher than expected ($p < 0.001$). These results suggest that 1) when three physicians share patients among themselves, they are all more likely to all be involved in risky prescribing than by chance; and 2) when two physicians share patients with a common third physician, and all three have been involved in risky prescribing, then these two physicians are more likely to also share patients between them than by chance. These results suggest that risky prescribing is in part driven by a higher-order (extra-dyadic) form of homophily involving ≥ 3 physicians, a more complex network phenomena than traditional dyadic homophily.

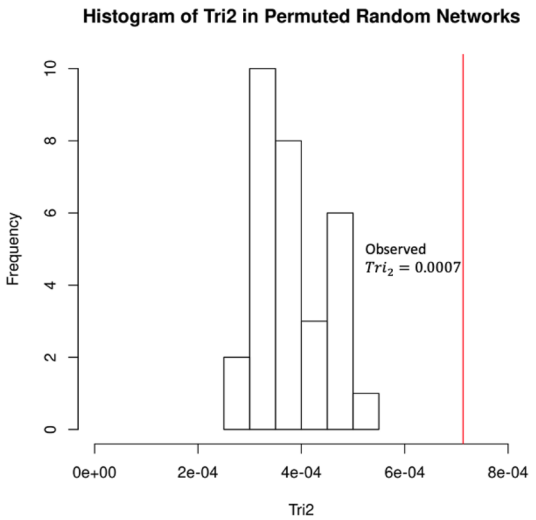
Discussion

This paper has made several methodological contributions to quantify physicians' prescribing and deprescribing behaviors comprehensively and to study homophily associated with prescribing in a shared-patient physician network. Using judiciously constructed algorithms we attributed physicians as responsible for observed changes in each patient's prescription drug status allowing each physician's contribution to their patients' prescribing and deprescribing transitions to be quantified in the form of a transition matrix. By summarizing these matrices in the form of multiple prescribing indexes and using ERGMs that included measures of the similarity of or distance between their prescribing indexes for each physician dyad, we developed a complete methodological procedure for estimating the independent effect of homophily of a physician prescribing index net of network density and the physicians' individual attributes (e.g., their speciality and the overall intensity of their prescribing).

Another methodological contribution is the development of two triadic homophily network statistics and associated statistical tests that avoided degeneracy in ERGMs. The two new network statistics and associated statistical tests advance the study of homophily from dyads to triads. Although we use standalone non-parametric attribute redistribution permutation tests to compare the observed statistics to those expected by chance, such statistics may be incorporated in ERGMs as network statistics so that their independent effect net of other network statistics included in the model can be estimated. A related direction of future research is to develop methods of generating the null distribution of the triadic homophily statistics considered herein while accounting for the effects of other network statistics and physician attributes.



(a) Triadic homophily statistic $Tri_1(\mathbf{a}, \mathbf{x})$



(b) Triadic homophily statistic $Tri_2(\mathbf{a}, \mathbf{x})$

Fig. 4 Histogram of triadic homophily network statistics generated by the non-parametric test for triadic homophily. The triadic homophily statistic $Tri_1(\mathbf{a}, \mathbf{x})$ is the proportion of closed triangles in the network in which each node has the $I_{everOBS}$ node attribute, reflecting whether a physician has ever contributed to bringing patients to the riskiest prescription state *OBS*. The triadic homophily statistic $Tri_2(\mathbf{a}, \mathbf{x})$ is the proportion of open two-paths with all nodes having the same attribute that are closed in the network. Panel **a** is the histogram of $Tri_1(\mathbf{a}, \mathbf{x})$ and panel **b** is the histogram of $Tri_2(\mathbf{a}, \mathbf{x})$ calculated from 30 networks with randomly shuffled node attributes under the null hypothesis of no homophily with respect to the given prescribing index. The red vertical lines denote the values in the observed network

In our risky prescribing application, we discovered substantial homophily of prescribing behaviors among physicians, as well as the assortative and disassortative mixing patterns with respect to physician specialty in the prescribing network. We also found significant risky-prescribing-associated homophily at the triadic level in the observed network compared to that expected by chance. We found that physicians' level of involvement in prescribing and deprescribing varied across specialties, and variation in prescribing-associated homophily across HRRs (see Supplemental Appendix for the

latter). Our findings of the homophily associated with prescribing behaviors and physicians' specialties in the shared-patient physician network provide a basis for promoting guideline-concordant prescribing practice and informing interventions (Ran et al. 2024a). The act of sharing patients can be a channel for behavior changes and so physicians who only share patients with risky prescribers might expose the focal physician to so much high-risk behavior that the prescribing practices of those physicians changes, forming a loop of problematic prescribing. Given this potential reinforcement of influence between physicians when homophily exists (Centola 2011), external interventions may be warranted to help break the cycle of risky prescribing among communities of guideline non-concordant prescribers.

Understanding the assortative and disassortative patterns among different physicians' specialties in the context of prescribing may provide an overarching view of the patient flow between different types of providers when patients seek medical care and prescriptions. The finding that prescribers in psychiatry were more likely to share patients with each other may be due to the complexity of conditions they generally encounter, requiring their patients to have multiple visits to different psychiatrists. In contrast, PCPs often refer patients to secondary care and are less likely to share patients among themselves. Given the variations in prescribing intensity across different physician specialties, our findings also suggest that it is worth targeting interventions to address risky prescribing. For example, if policymakers were to impose guidelines to promote safe prescribing among different physician specialties, they may expect differential impacts as specialties do not all have the same baseline prescribing intensities and some of the specialties (e.g., psychiatry) may have better outcomes of interventions because they are more likely to share patients with physicians in the same specialty than across specialties.

Various aspects of the methodology developed in this paper can be adapted or generalized to study the homophily of prescribing behaviors for other classes of drugs or medical procedures, especially if they may be repeatedly performed across time and their use may be considered discretionary (e.g., harmful or guideline inconsistent in some situations). For instance, the methods developed herein may be used to construct a physician network and indexes of physician treatment appropriateness; the later could involve matrices of transitions in patient treatment status that each physician was responsible for or a different data object. ERGMs or another network model may be used to determine whether the professional relationships between the physicians treating the patients receiving the given treatments or procedures are clustered (are more likely to exist) among physicians with the same or similar treatment tendencies. Furthermore, if there is interest in knowing whether homophily occurs at least in-part as a higher-order phenomenon than a purely dyadic phenomenon, tests of extra-dyadic homophily like those constructed herein may be used. The methodology also extends beyond medicine to any situation in which the goal is to determine whether homophily and extra-dyadic homophily of an actor attribute are associated with the structure of a network (e.g., between risk-taking behavior of traders on stock exchanges or other financial markets). All that is needed is a network and a node attribute of interest x.

This study is subject to several limitations. First, although our study is the first to our knowledge to deduce deprescribing from administrative claims data, not having deprescribing events recorded let alone time-stamped in claims data meant that we needed to

define deprescribing heuristically, likely resulting in the attribution of deprescribing not being as sensitive as ideal (as evinced by the distributions of I_0 and I_1 being dominated by prescribing events). The availability of Electronic Health Record (EHR) data might have enabled deprescribing to be more accurately identified. However, despite the limited data we had, the involvement of deprescribing in our indexes was critical as it allowed measures of the net excess of prescribing over deprescribing to be used as indexes. Second, the data used in this study was cross-sectional, which led to challenges in estimating network effects beyond homophily. The availability of longitudinal data would have allowed dyadic dependent network effects to be modeled as lagged variables to avoid degeneracy (Paul and O'Ḿalley 2013). Longitudinal data would also have allowed the process of social selection (i.e., the factors governing the selection of relationships (Runciman 2009)) to be more easily distinguished from social influence (i.e., the process of one individual exerting influence on another so that they adopt similar traits). Third, our study focused on the Medicare population, whereas the same research question within younger populations is also of interest.

Conclusion

In conclusion, we proposed a novel framework to model the relationship between informal physician professional networks and new measures for quantifying physicians' prescribing and deprescribing behavior. We discovered significant homophily and extra-dyadic homophily associated with prescribing among physicians' connections through sharing patients. These findings provide important insights into the mechanism underlying the spread of risky prescribing among the older population in the United States and of how communities of prescribers emerge and evolve. We hope that this work helps to incentivize interventions to reduce practices that are not compliant with guidelines and to promote safe practices among healthcare providers.

Supplemental appendix

The supplemental appendix contains sections describing the identification of prescribing and deprescribing events and their attribution to one or more physicians. It also includes a visualization of the network, a goodness-of-fit analysis of the fit of one of the estimated ERGMs and analyses of the heterogeneity in the direction and level of homophily across health referral regions in the United States. These materials are intended to be published with the paper and will also be available at the paper's GitHub site listed above.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s41109-024-00670-y.

Author contributions

All authors reviewed and approved the submission of the manuscript. In addition: X.R. conducted the computer programming, implemented the methods, and drafted a substantial part of the manuscript. E.M. provided detailed guidance on the analysis of claims data and interpreted the results of the analysis of our motivating application. N.M. gave expert guidance on data wrangling the Part D Medicare drug claims and helped to interpret the results of the analysis of our motivating application. E.M. gave critical review on study design and the framing of the manuscript. D.R. critically reviewed the quantitative methods and reviewed the manuscript. A.J.O. conceptualized the research, developed the methodological ideas, wrote some of the initial manuscript, revised the manuscript text following peer-review, and was responsible for ensuring that the research was conducted appropriately and for checking the accuracy of the entire manuscript.

Funding
The research in this paper was supported by grant P01 AG019783 from the National Institutes of Health (NIH).

Availability of data and materials
The data used for the motivating analyses contains patient-identifiable information and so cannot be made available. However, R code for performing simulation studies (which can be easily adapted to analyze a real data set) is available from the paper's GitHub site: https://github.com/xinran02/PolyRxNetworkHomophily

Declarations

Competing interests
The authors declare no potential Conflict of interest.

Received: 20 March 2024 Accepted: 28 August 2024
Published online: 03 October 2024

References

An C, O'malley AJ, Rockmore DN et al (2018) Analysis of the US patient referral network. Stat Med 37(5):847–866
An C, O'Malley AJ, Rockmore DN (2018) Referral paths in the US physician network. Appl Netw Sci 3(1):1–24
Apicella CL, Marlowe FW, Fowler JH et al (2012) Social networks and cooperation in hunter-gatherers. Nature 481(7382):497–501
Aroke H, Katenka N, Kogut S, et al (2021) Network-based analysis of prescription opioids dispensing using exponential random graph models (ergms). In: International Conference on Complex Networks and Their Applications, Springer, pp 716–730
Barnett ML, Christakis NA, O'Malley AJ et al (2012) Physician patient-sharing networks and the cost and intensity of care in us hospitals. Med Care 50(2):152
Bushardt RL, Massey EB, Simpson TW et al (2008) Polypharmacy: misleading, but manageable. Clin Interv Aging 3(2):383–389
Centers for Disease Control and Prevention and others (2016) Guideline for prescribing opioids for chronic pain. J Pain Palliative care Pharmacother 30(2):138–140
Centers for Medicare & Medicaid Services (2012) National plan and provider enumeration system. https://download.cms.gov/nppes/NPI_Files.html
Centola D (2011) An experimental study of homophily in the adoption of health behavior. Science 334(6060):1269–1272
Cho J, Spence MM, Niu F et al (2020) Risk of overdose with exposure to prescription opioids, benzodiazepines, and non-benzodiazepine sedative-hypnotics in adults: a retrospective cohort study. J Gen Intern Med 35(3):696–703
Contractor NS, Wasserman S, Faust K (2006) Testing multitheoretical, multilevel hypotheses about organizational networks: an analytic framework and empirical example. Acad Manag Rev 31(3):681–703
Dreischulte T, Guthrie B (2012) High-risk prescribing and monitoring in primary care: how common is it, and how can it be improved? Ther Adv Drug Saf 3:175–184
Farrell B, Mangin D (2019) Deprescribing is an essential part of good prescribing. Am Fam Phys 99(1):7–9
Fattore G, Frosini F, Salvatore D et al (2009) Social network analysis in primary care: the impact of interactions on prescribing behaviour. Health Policy 92(2–3):141–148
Fulton MM, Riley Allen E (2005) Polypharmacy in the elderly: a literature review. J Am Acad Nurse Pract 17(4):123–132
Gnjidic D, Le Couteur DG, Pearson SA et al (2013) High risk prescribing in older adults: prevalence, clinical and economic implications and potential for intervention at the population level. BMC Public Health 13(1):1–9
Goodreau SM (2007) Advances in exponential random graph (p*) models applied to a large social network. Soc Netw 29(2):231–248
Handcock MS (2003) Assessing degeneracy in statistical models of social networks. J Am Stat Assoc 76:33–50
Handcock MS, Hunter DR, Butts CT et al (2008) statnet: software tools for the representation, visualization, analysis and simulation of network data. J Stat Softw 24(1):1548
Hunter DR, Handcock MS, Butts CT et al (2008) ergm: A package to fit simulate and diagnose exponential-family models for networks. J Stat softw 24(3):1–29. https://doi.org/10.18637/jss.v024.i03
Landon BE, Keating NL, Barnett ML et al (2012) Variation in patient-sharing networks of physicians across the united states. JAMA 308(3):265–273
Latapy M, Magnien C, Del Vecchio N (2008) Basic notions for the analysis of large two-mode networks. Soc Netw 30(1):31–48
Levy B, Paulozzi L, Mack KA et al (2015) Trends in opioid analgesic-prescribing rates by specialty, us, 2007–2012. Am J Prev Med 49(3):409–413
Mascia D, Di Vincenzo F, Iacopino V et al (2015) Unfolding similarity in interphysician networks: the impact of institutional and professional homophily. BMC Health Serv Res 15(1):1–8
McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Ann Rev Sociol 27(1):415–44
Moen EL, Austin AM, Bynum JP et al (2016) An analysis of patient-sharing physician networks and implantable cardioverter defibrillator therapy. Health Serv Outcomes Res Method 16(3):132–153
Morris M, Handcock MS, Hunter DR (2008) Specification of exponential-family random graph models: terms and computational aspects. J Stat Softw 24(4):1548

O'Malley AJ, Marsden PV (2008) The analysis of social networks. Health Serv Outcomes Res Method 8(4):222–269. https://doi.org/10.1007/s10742-008-0041-z
O'Malley AJ, Ran X, An C et al (2022) Optimal physician shared-patient networks and the diffusion of medical technologies. J Data Sci 21(3):578–598
Paul S, O'Malley AJ (2013) Hierarchical longitudinal models of relationships in social networks. J R Stat Soc Ser C (Appl Stat) 62(5):705–722
Pollack CE, Soulos PR, Gross CP (2015) Physician's peer exposure and the adoption of a new cancer treatment modality. Cancer 121(16):2799–2807
Quinn KJ, Shah NH (2017) A dataset quantifying polypharmacy in the United States. Sci data 4(1):1–11
R Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/
Ran X, Meara ER, Morden NE, et al (2024a) Modeling the association between physician risky-prescribing and the dyadic and network structure of physician shared-patient relationships. In: Cherifi H, Rocha L, Cherifi C, et al (eds) Complex Networks & Their Applications XII. COMPLEX NETWORKS 2023, vol 1143. Springer, Cham, https://doi.org/10.1007/978-3-031-53472-0_17
Ran X, Morden NE, Meara ER, et al (2024b) Modeling the association between physician risky-prescribing and the complex network structure of physician shared-patient relationships. In Press: Statistics in Medicine Res Sq [Preprint]. 2024 Mar 26:rs.3.rs-4139630https://doi.org/10.21203/rs.3.rs-4139630/v1
Rand DG, Arbesman S, Christakis NA (2011) Dynamic social networks promote cooperation in experiments with humans. Proc Natl Acad Sci 108(48):19193–19198
Robins G, Pattison P, Kalish Y et al (2007) An introduction to exponential random graph (p*) models for social networks. Soc Netw 29(2):173–191
Robins G, Snijders T, Wang P et al (2007) Recent developments in exponential random graph (p*) models for social networks. Soc Netw 29(2):192–215
Runciman WG et al (2009) The theory of cultural and social selection. Cambridge University Press, New York
Snijders TA, Pattison PE, Robins GL et al (2006) New specifications for exponential random graph models. Soc Methodol 36(1):99–153
Sun EC, Dixit A, Humphreys K et al (2017) Association between concurrent use of prescription opioids and benzodiazepines and overdose: retrospective analysis. Bmj 356:j760
Van Rossum G, Drake FL (2009) Python 3 Reference Manual. CreateSpace, Scotts Valley
Werder SF, Preskorn SH (2003) Managing polypharmacy: walking the fine line between help and harm. Curr Psychiatry Online 2(2):24–36

Publisher's Note