Causal Discovery of Urban Design Factors Associated with Public Sentiment Using Geospatial Machine Learning

Viveka Blomgren
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
blomgren1987@binghamton.edu

Jodi Conway
School of Computing, Clemson University, Clemson, SC, USA.
jodimail@clemson.edu

Abstract

Urban design features such as street greening, enclosure, walkability, and facade complexity are hypothesized to shape how people feel in public space, yet most geospatial machine-learning studies optimize predictive accuracy rather than identify causal relations. This paper develops a causal discovery framework for estimating which urban design factors are associated with public sentiment under explicit identification assumptions. The framework treats street segments or grid cells as units, constructs design exposures from street view imagery and geographic information systems, derives sentiment from geotagged text and visual context, and learns a constrained causal graph over design, contextual, spatial, and sentiment variables. Continuous structure learning with temporal and domain restrictions separates design effects from spatial confounding, residential sorting, and platform selection, while a spatial autoregressive confounder model represents unobserved neighborhood dependence. Effects are standardized through g-computation, and uncertainty is quantified with spatial block bootstrap and graph-stability diagnostics. The contribution is not a new sentiment classifier but an end-to-end causal estimand and inference design that links multimodal geospatial measurement to intervention-oriented questions in urban design. The analysis shows why predictive importance, partial correlations, and ordinary spatial regression can misstate design effects when sentiment is selectively reported and design features co-vary at multiple scales. It also specifies boundary conditions for interpretation, including positivity, interference, measurement invariance, and graph faithfulness. The framework provides a reproducible basis for evaluating design scenarios, prioritizing street-level interventions, and reporting causal claims with appropriate spatial and measurement uncertainty.

Keywords

causal discovery; urban design; public sentiment; geospatial machine learning; spatial confounding; street view imagery

1. Introduction

Urban design is an intervention science in which street-level decisions about vegetation, enclosure, sidewalk width, facade articulation, and intersection configuration are expected to alter experience, behavior, and social interaction. The rise of geotagged text, street view imagery, and urban sensor data has made it possible to measure public sentiment at fine spatial resolution, but measurement abundance does not by itself yield causal knowledge.

Causal claims require a model of the data-generating process, not only a predictive fit [1], and graph-based identification clarifies when adjustment sets are adequate [2]. Continuous structure learning makes directed acyclic graph search differentiable [3], and structural causal models provide the language for expressing interventions rather than passive associations [4]. These developments matter for urban design because the same design feature may appear important in a predictive model while being confounded by neighborhood composition, historical investment, spatial dependence, or selective reporting of sentiment.

Built-environment research has long measured density, diversity, and design as correlates of travel and activity [5], [6], and later work has refined the measurement of urban design qualities that are not captured by coarse land-use indicators [7]. Sentiment geography links expressed affect to place characteristics [8], while volunteered geographic information provides high-volume observations of human activity and perception [9]. Big data and smart-city analytics have expanded the observational base for urban planning [10], but they have also encouraged a substitution of predictive performance for causal identification. A model that predicts sentiment well from streetscape features can still be biased for intervention because it does not represent what would happen if a street were redesigned, and because spatially clustered features are entangled with unobserved neighborhood conditions.

This paper addresses that gap by specifying a causal discovery framework for urban design factors associated with public sentiment using geospatial machine learning. The central contribution is a coherent estimand-and-inference design that links multimodal measurement, constrained graph learning, spatial confounding adjustment, and intervention standardization. The framework is not an end-to-end predictor that treats sentiment as a label to be forecast from imagery. It is a structured analysis in which design exposures are defined as potential interventions, sentiment is treated as a measured outcome with explicit error, spatial dependence is modeled rather than residualized away, and graph restrictions encode temporal and design-knowledge constraints. The discussion explains why credible causal claims require boundary conditions on interference, positivity, measurement invariance, and graph faithfulness, and how those conditions can be reported with spatial uncertainty.

2. Problem Formulation and Causal Target

Let i = 1,…,n index street segments or grid cells, and let t = 1,…,T_i index time windows in which sentiment observations are available. The design exposure vector is D_i ∈ R^p, with components such as visible greenery, sky view, enclosure, sidewalk coverage, street width, building height, facade complexity, and intersection density. The observed context vector is C_i, containing land use, population density, income, transit access, and other compositional variables that may affect both design and sentiment. The latent spatial confounder is U_i, representing unobserved neighborhood conditions such as maintenance, perceived safety, historical investment, or social capital. The structural outcome model is written as

$$S_i t = f(D_i, C_i, U_i, t) + \varepsilon_i t$$

where S_it is latent public sentiment for unit i in period t, f is an unknown structural function, and ε_it is an idiosyncratic error. The target is not the predictive decomposition of S_it but the consequence of intervening on a design component. For the j-th design feature, with observed value d_j and counterfactual value d′_j, the average causal effect is

$$\tau_j \;=\; E[S_i \mid do(D_ij = d'_j), C_i] \;-\; E[S_i \mid do(D_ij = d_j), C_i]$$

where do(·) denotes an intervention that sets the feature while leaving other design and contextual variables to their natural values under the stated intervention policy. The estimand is defined at the scale of the spatial unit and must be interpreted with respect to the intervention, the time window, and the neighborhood exposure mapping. Identification requires consistency, positivity for the design values compared, and no unmeasured design-sentiment confounding conditional on C_i and the modeled spatial structure. Because street redesign can affect adjacent units, the stable-unit treatment-value assumption is replaced by an explicit neighborhood exposure mapping that includes spillovers within a defined buffer.

Spatial dependence is not a nuisance to be removed mechanically. Local indicators of spatial association reveal clustering and spatial outliers that can reflect shared causes of design and sentiment [11], and spatial econometric models formalize dependence through lagged or error structures [12]. The causal graph therefore includes both observed context and latent spatial dependence. Domain restrictions orient edges according to temporal order and design theory: sentiment cannot cause the existing street width, and a current sentiment observation cannot cause a past land-use pattern. These restrictions reduce Markov equivalence classes and prevent the discovery algorithm from proposing substantively impossible directions, while still allowing the data to determine which design features are adjacent to sentiment conditional on context.

3. Geospatial Data and Feature Representation

The measurement layer combines three sources: geotagged text and reviews for expressed sentiment, street view imagery for visible streetscape qualities, and geographic information systems for structural and land-use attributes. Text-derived sentiment can be represented through topic models such as latent Dirichlet allocation [13] or through contextual language encoders such as bidirectional transformer models [14]. The choice affects both construct validity and error structure: lexicon scores are transparent but brittle to negation and local slang, whereas contextual encoders capture more semantic nuance but require annotated calibration and attention to domain shift. Street view imagery supports deep-learning models that quantify perceived qualities at scale [15], and combining geotagged posts with street view imagery and deep neural networks enables streetscape modeling across large areas [16]. Street view audits have also been used to characterize neighborhood environments for health and planning research [17], and human-machine scoring frameworks show how visual perception can be elicited with structured comparison tasks [18].

The sentiment measurement model is specified as

$$S*_it \;=\; g_\theta(T_it, V_i) \;+\; \eta_it$$

where S_it is the measured sentiment score, T_it is the text associated with unit i and period t, V_i is the visual context, g_θ is a calibrated sentiment function with parameters θ, and η_it is measurement error. The model does not assume that S_it equals S_it. Instead, it treats sentiment as a latent construct observed through platform-specific expression, and it requires calibration against held-out human annotations, tests of measurement invariance across neighborhoods, and sensitivity to platform composition. A design feature vector is aggregated from sampled street view points P_i and image or GIS attributes as

$$D_ik \;=\; (1/|P_i|)\,\sum_\{p \in P_i\}\, h_k(I_ip)$$

where D_ik is the k-th design feature for unit i, I_ip is the image or image-derived attribute at point p, and h_k is the feature model for that design dimension. The sampling design of P_i matters: points must cover both sides of a street, avoid systematic occlusion,

and represent the pedestrian rather than vehicle viewpoint. Image-derived greenery, enclosure, and facade complexity should be validated against field audits or manually scored subsets before they are used as causal exposures.

Geotagged sentiment is selectively produced. Reporting intensity depends on platform use, event attendance, commercial activity, and affective arousal, so the probability of observing a sentiment post is not independent of the outcome. This selection process can be represented as a reporting kernel that depends on sentiment, context, and design, and it motivates weighting, negative-control outcomes, and sensitivity analysis rather than an assumption that missing posts are missing completely at random. Spatial coverage also varies: dense commercial districts generate more posts than residential alleys, and street view capture dates may not match the sentiment observation window. The feature representation must therefore carry provenance, capture date, and spatial-support metadata so that causal estimates are not contaminated by temporal mismatch or by systematic differences in image availability.

4. Causal Discovery Framework

The causal discovery problem is formulated over a variable system X that includes design features, observed context, spatial summaries, and measured sentiment. The graph is represented by a weighted adjacency matrix B, and unit-specific structural parameters are represented by Θ_i to allow effect heterogeneity across neighborhoods. The learning objective combines a score or likelihood term with sparsity, spatial smoothing, and an acyclicity penalty:

$$L(B,\Theta) = \ell(B,\Theta; X) + \lambda 1||B||_1 + \lambda 2 \sum_i \sum_\{j \in N(i)\} ||\Theta_i - \Theta_j||_2 + \lambda 3\, h(B)$$

where ℓ(B,Θ; X) is a score or likelihood term, λ1, λ2, and λ3 are regularization strengths, N(i) is the set of spatial neighbors of unit i, and h(B) is an acyclicity function. A standard continuous acyclicity penalty is h(B) = tr(exp(B◦ B)) − d, where d is the number of variables and ◦ denotes the Hadamard product. The L1 term enforces sparse edges, the spatial term discourages implausible abrupt changes in structural coefficients between adjacent units, and the acyclicity term enforces a directed acyclic graph during continuous optimization.

The score term can be a Gaussian likelihood when variables are continuous and linear relations are adequate, but it can also be a flexible conditional likelihood or a cross-fitted machine-learning score. The graph is constrained so that measured sentiment has no outgoing edges to design or context, and so that land-use and population variables can precede sentiment but cannot be caused by current sentiment. Temporal ordering across repeated observations provides additional orientation constraints. The spatial smoothing term should not be interpreted as evidence that causal effects are spatially homogeneous; it is a regularizer that trades bias for stability when units are sparse. Sensitivity analysis should compare graphs learned with different spatial smoothing strengths, because excessive smoothing can erase real street-level heterogeneity and too little smoothing can fit local noise.

Causal discovery under spatial dependence is not equivalent to applying a generic DAG learner to georeferenced rows. Rows are not independent, and neighborhood similarity can create edges that reflect shared unobserved causes rather than direct effects. Nonlinear dependence tests and causal association measures can complement graph search when associations are not linear or Gaussian [19], but they do not remove the need for spatial adjustment. The graph should therefore be evaluated with spatial block cross-validation, edge-stability diagnostics across bootstrap samples, and domain review of every proposed edge involving a design exposure. The output is not a single accepted graph but a set of plausible

structures whose implications for the target effect are examined through standardization and sensitivity analysis.

5. Estimation, Identification, and Uncertainty

After graph restrictions are imposed, the target effect is estimated by standardization rather than by reading a regression coefficient. Let m_j(d, C_i) denote the expected sentiment under setting the j-th design feature to value d and conditioning on context C_i. The g-computation estimator is

$$\tau_j \;=\; (1/n)\sum_\{i = 1\}\^n\, [m_j(d'_j, C_i) \;-\; m_j(d_j, C_i)]$$

where n is the number of spatial units and the empirical distribution of C_i approximates the target population of contexts. The estimator is valid under the identification assumptions stated above, including positivity and no unmeasured confounding conditional on the adjustment set and spatial structure. If the outcome model is misspecified, the estimator can be biased; cross-fitting with flexible learners can reduce model dependence, but it does not substitute for identification. The same caution applies to matching or weighting estimators: they change the estimation strategy, not the causal assumptions.

Unobserved spatial confounding is represented through a latent autoregressive structure:

$$U_i = \rho \sum_{j \in N(i)} w_i j U_j + \nu_i$$

where U_i is the latent neighborhood confounder, w_ij are spatial weights, ρ is a dependence parameter, and ν_i is an independent disturbance. This expression clarifies why spatially lagged outcomes or spatially autoregressive errors can partly absorb confounding, but it also shows the limitation: if U_i is unobserved, the parameter is not directly identified without additional structure. The practical response is to estimate spatial random effects or spatial error models, report sensitivity to the assumed spatial weights, and use negative-control outcomes that should not be affected by the design intervention. A design factor that changes sentiment but also changes a negative-control outcome in the same direction is a warning that the estimated effect may reflect a shared spatial cause.

Uncertainty should be computed at the level of the spatial unit rather than the individual sentiment post, because posts within a unit are not independent and because the unit is the policy-relevant scale. Spatial block bootstrap resamples contiguous neighborhoods, preserving local dependence, and the variance is

$$Var*(\tau_j) \;=\; (1/B)\sum_\{b = 1\}\^B\, (\tau_j\^(b) \;-\; \bar{\tau}_j)\^2$$

where B is the number of bootstrap replicates, τ_j^(b) is the estimated effect in replicate b, and $\bar{\tau}$_j is the mean across replicates. Graph uncertainty is reported separately through edge frequency and effect-range summaries across plausible graphs. This separation matters: a narrow confidence interval conditional on one graph can understate total uncertainty when the graph itself is unstable.

6. Analytical Assessment and Technical Discussion

The proposed framework differs from predictive importance analysis in several consequential ways. A random forest or gradient-boosted model can rank design features by predictive contribution, but such rankings are sensitive to correlated features, spatial autocorrelation, and the inclusion of proxies for unobserved neighborhood quality. A feature may appear important because it substitutes for income, maintenance, or commercial activity

rather than because it is a plausible intervention target. The causal graph approach forces the analyst to state which variables are causes, which are context, which are consequences of sentiment, and which are unobserved. This does not guarantee truth, but it makes the assumptions inspectable and the estimand explicit.

Ordinary spatial regression is also insufficient when the goal is intervention. A spatially lagged outcome model can improve prediction by borrowing strength from neighbors, yet the coefficient on a design feature may still be confounded by common causes at the neighborhood scale. The graph-based design includes spatial dependence as part of the data-generating process and then standardizes over observed context. The trade-off is that spatial smoothing and graph constraints introduce regularization bias. That bias is acceptable only when it is reported and when sensitivity analyses show that the substantive conclusion does not depend on a narrow choice of spatial weights or penalty strengths.

Measurement error in sentiment is a first-order concern. Text sentiment is not equivalent to experienced well-being, and a positive post about a restaurant may not indicate that the street is pleasant. Visual streetscape features are also measured through image models whose errors may correlate with neighborhood upkeep, weather, and camera angle. The framework therefore treats image and text outputs as noisy measurements of latent constructs and requires validation against human audits, temporal stability checks, and falsification tests. A defensible claim is conditional: under a specified measurement model and adjustment set, changing visible greenery or enclosure by a stated amount is associated with a standardized shift in sentiment for the target units. The result should not be generalized to a different city, platform population, or design scale without rechecking positivity and measurement invariance.

7. Implications and Boundary Conditions

For urban design practice, the framework supports scenario analysis rather than single-number feature importance. A planner can compare a greening intervention, a sidewalk-widening intervention, and a facade-articulation intervention using the same standardization logic, while reporting uncertainty and the spatial distribution of effects. The analysis can identify where a design factor is well supported by local variation and where positivity is weak because almost all streets in a district share the same configuration. It can also reveal whether sentiment effects are concentrated in particular contexts, such as high-density commercial streets or low-density residential blocks, without treating those contextual differences as nuisance heterogeneity.

The boundary conditions are substantial. Interference can invalidate unit-level estimates when a redesigned street changes pedestrian flows or commercial activity in adjacent streets. The modifiable areal unit problem affects any grid or segment definition, so results should be reported at more than one spatial scale where data permit. Sentiment platforms overrepresent some populations and activities, and street view imagery may be dated relative to the sentiment observation window. Causal graph discovery assumes faithfulness and sufficient adjustment, and those assumptions are not guaranteed in observational urban data. These limitations do not make the framework unusable; they define the claims that can be defended and the sensitivity analyses that must accompany them.

8. Conclusion

This paper has developed a causal discovery framework for urban design factors associated with public sentiment using geospatial machine learning. The framework defines

design exposures as potential interventions, treats sentiment as a noisy measured outcome, learns a constrained graph over design, context, spatial dependence, and sentiment, and estimates effects through g-computation with spatial block uncertainty. Its central methodological claim is that multimodal geospatial data can support causal questions only when predictive measurement is joined to explicit identification, spatial confounding adjustment, and reporting of boundary conditions. The framework provides a disciplined alternative to feature-importance rankings and unadjusted spatial regression, and it makes the link between street-level design scenarios and sentiment outcomes analytically transparent.

References

1. Pearl, J. (2009). Causality: Models, reasoning, and inference (2nd ed.). Cambridge University Press.
2. Spirtes, P., Glymour, C., & Scheines, R. (2000). Causation, prediction, and search (2nd ed.). MIT Press.
3. Zheng, X., Aragam, B., Ravikumar, P., & Xing, E. P. (2018). DAGs with NO TEARS: Continuous optimization for structure learning. In Advances in Neural Information Processing Systems (Vol. 31, pp. 9472–9483). Curran Associates.
4. Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of causal inference: Foundations and learning algorithms. MIT Press.
5. Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. Journal of the American Planning Association, 76(3), 265–294.
6. Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. Transportation Research Part D: Transport and Environment, 2(3), 199–219.
7. Ewing, R., Handy, S., Brownson, R. C., Clemente, O., & Winston, E. (2006). Identifying and measuring urban design qualities related to walkability. Journal of Physical Activity and Health, 3(S1), S223–S240.
8. Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. PLoS ONE, 8(5), e64417.
9. Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. GeoJournal, 69(4), 211–221.
10. Batty, M. (2013). Big data, smart cities and city planning. Dialogues in Human Geography, 3(3), 274–279.
11. Anselin, L. (1995). Local indicators of spatial association—LISA. Geographical Analysis, 27(2), 93–115.
12. LeSage, J., & Pace, R. K. (2009). Introduction to spatial econometrics. CRC Press.
13. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3, 993–1022.
14. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational

Linguistics: Human Language Technologies (pp. 4171–4186). Association for Computational Linguistics.

15. Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. In Computer Vision – ECCV 2016 (pp. 196–212). Springer.
16. Aman, J., Matisziw, T. C., Kim, J. B., & Luo, D. (2022, April). Sensing the city: Leveraging geotagged social media posts and street view imagery to model urban streetscapes using deep neural networks. In Proceedings of the International Conference on Computer-Aided Architectural Design Research in Asia (pp. 595-604). Association for Computer-Aided Architectural Design Research in Asia.
17. Rundle, A. G., Bader, M. D. M., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using Google Street View to audit neighborhood environments. American Journal of Preventive Medicine, 40(1), 94–100.
18. Yao, Y., Liang, Z., Yuan, Z., Liu, P., Bie, Y., Zhang, X., Wang, R., & Guan, Q. (2019). A human-machine adversarial scoring framework for urban perception assessment using street-view images. International Journal of Geographical Information Science, 33(12), 2363–2384.
19. Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. Science Advances, 5(11), eaau4996.