

# Measuring Housing Quality Using Revealed Preference: A Geographic PageRank Approach

---

CRIW Conference  
on Measurement of Housing and the Housing Sector  
March 12, 2026

Alex Bell (Georgia State University)  
Sophie Calder-Wang (University of Pennsylvania)  
Shusheng Zhong (Northwestern University)

# Introduction

- **Research Question:**

- **How can one measure the value of a place for households?**

- The value of a place is reflected through
      - the cost of housing (observed) + the value of amenities (not directly observed)

- **Approach:**

1. *We propose a novel measure called Geographic PageRank (GPR)*

- A ranking of places revealed by household migration choices
  - A measure that captures a network-based measure of centrality
    - Places attracting people are ranked as good
    - Places attracting people from other attractive places are ranked as even better
  - Interpreted via a revealed preference framework as households generally try to move to better places

2. *Apply the algorithm to rich data to obtain rankings*

- Across geography and time periods
  - Across sub-populations

3. *Use the Geographic PageRank as a measure of unobserved housing quality*

- We show how to use it to pricing amenities (e.g., air quality) using observational data

# Outline

- I. Motivation
- II. Theory of the PageRank Algorithm
- III. Empirics and Measurement
- IV. Application for pricing amenities

# The PageRank Algorithm

## (12) **United States Patent** **Page**

---

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) Inventor: **Lawrence Page, Stanford, CA (US)**

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University, Stanford, CA (US)**

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/004,827**

(22) Filed: **Jan. 9, 1998**



## (57) **ABSTRACT**

A method assigns **importance ranks** to nodes in a linked database, such as any database of documents containing citations, **the world wide web** or any other hypermedia database. **The rank assigned to a document is calculated from the ranks of documents citing it.** In addition, the rank of a document is calculated from a constant representing the probability that a browser through the database will randomly jump to the document. The method is particularly useful in enhancing the performance of search engine results for hypermedia databases, such as the world wide web, whose documents have a large variation in quality.

- PageRank, at its core, is an **iterative** eigenvector computation to produce a measure of network centrality
- Additional features added to resist manipulation (e.g., damping factor)

# The PageRank Algorithm: A Brief History

## (12) **United States Patent Page**

---

(54) **METHOD FOR NODE RANKING IN A LINKED DATABASE**

(75) Inventor: **Lawrence Page**, Stanford, CA (US)

(73) Assignee: **The Board of Trustees of the Leland Stanford Junior University**, Stanford, CA (US)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/004,827**

(22) Filed: **Jan. 9, 1998**

- The Eigenvalue problem was discovered and rediscovered:
  - In 1895, Edmund Landau (mathematician) suggested for ranking chess players
  - In 1976, Gabriel Pinski and Francis Narin used it to rank scientific journals
  - ...
  - In 1996, Robin Li patented RankDex and founded Baidu
  - In 2001, Larry Page patented PageRank
  - Nowadays, it is used in a wide variety of settings
    - Bibliometrics, network analysis, link prediction and recommendation
    - Biology, chemistry, neuroscience, and physics
    - Sorkin (2018): Ranking Firms Using Revealed Preference



# Related Literature

- **PageRank:**

- Web: Brin and Page (1998), Page et. al. (1999), Page (2001)
- Numerous applications across many fields(survey by Gleich 2015)
  - Including a ranking of roads in transportation studies (Jiang 2009)
- Labor economics:
  - Firm rankings: Sorkin (2018), Lachowska, Mas, Saagio, and Woodbury (2023); Morchio and Moser (2024)

- **Housing:**

- Modeling housing markets through a single index of quality:
  - Epple, Quintero, and Sieg (2020), Landvoigt, Piazzesi, and Schneider (2015), Ekeland, Heckman, Nesheim (2004)
- Role of amenities in household location choices:
  - Diamond (2016), Almagro, Dominguez-lino (2024)

# Define Geographic PageRank (GPR)

- Geographic PageRank:

- **Adjacency matrix  $M$ :**

- $M_{ij}$  represents the fraction of households migrating from location  $j$  to  $i$

- **Eigenvector formulation:**

$$Mv = v$$

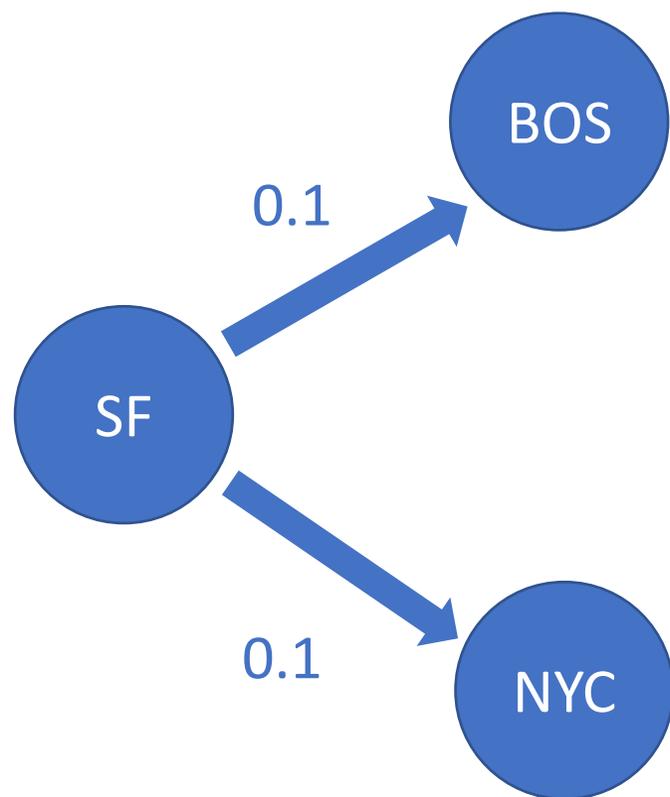
- $v$  is the stationary distribution of the Markov process  $M$ 
      - If people keep migrating with the migration matrix  $M$  forever,  $v$  represents the eventual population distribution that it converges to
      - Places with higher probabilities are ranked higher

- **Geographic PageRank formulation**

$$(dM + (1-d)U)v = v$$

- The damping factor  $d$  ( $=0.85$ ) and  $U$  represent a uniform random teleport
    - Better computational properties (connected, no cycling) and robust to manipulation (link farms)

# PageRank Example



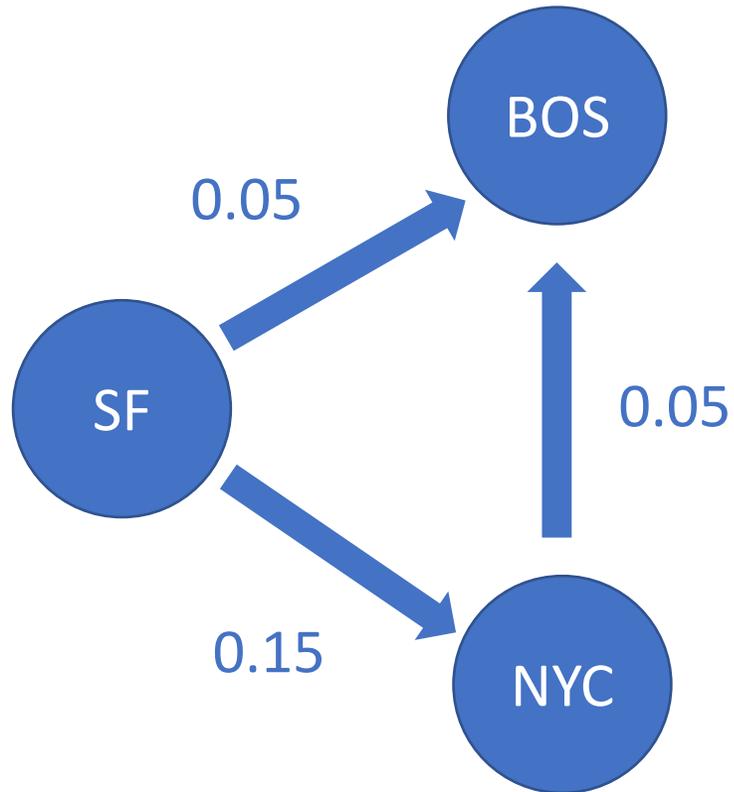
## Starting Input:

- Equal starting size
- 10% leaves SF for Boston
- 10% leaves SF for NYC

## Output:

- Net migration rates
  - SF: -20%; NYC: +10%, BOS: +10%
- Ranking:
  - Boston = NYC > SF

# PageRank Example



## Starting Input:

- Equal starting size
- 5% leaves SF for Boston
- 15% leaves SF for NYC
- 5% leaves NYC for Boston

## Output:

- Same net migration rates as before
  - SF: -20%; NYC: +10%, BOS: +10%
- Ranking
  - Boston > NYC > SF

# Outline

- I. Motivation
- II. Theory of the PageRank Algorithm
- III. Empirics and Measurement**
- IV. Application for pricing amenities

# Overview of Empirics

- **Data:**

1. IRS County-to-county migration counts
  - Available from 1991 to present
  - Cover the entire universe of population with Form 1040 filings
2. ACS micro data
  - Available from 2005 to present
  - Rich household and individual demographics
  - Migration PUMA is created for data privacy, not consistent over time
    - Focus is on metropolitan level rankings to compare ranks for different demographic groups
3. Potential sources for future users of the algorithm
  - Restricted census with finer geographic granularities
  - Marketing-data with address histories (e.g., Data Axle, Infutor)

# Outline

- I. Motivation
- II. Theory of the PageRank Algorithm
- III. Empirics and Measurement
  - a) Cross-sectional rankings
  - b) Time-series rankings
  - c) Rankings by demographic subgroups
- IV. Application for pricing amenities

# Geographic PageRank for US Counties

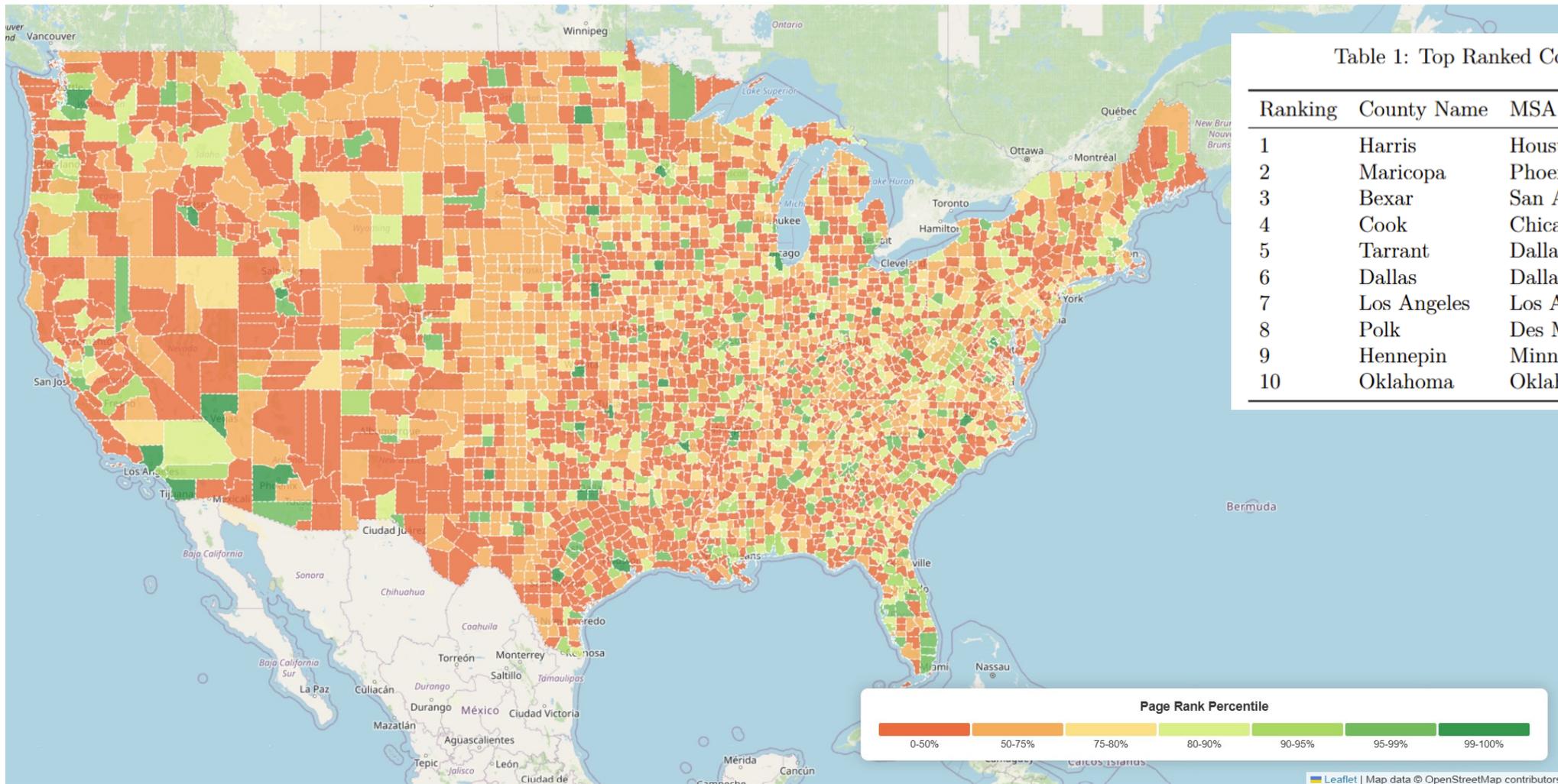


Table 1: Top Ranked Counties in the U.S. (2021-2022)

Ranking	County Name	MSA Name
1	Harris	Houston-Pasadena-The Woodlands, TX
2	Maricopa	Phoenix-Mesa-Chandler, AZ
3	Bexar	San Antonio-New Braunfels, TX
4	Cook	Chicago-Naperville-Elgin, IL-IN
5	Tarrant	Dallas-Fort Worth-Arlington, TX
6	Dallas	Dallas-Fort Worth-Arlington, TX
7	Los Angeles	Los Angeles-Long Beach-Anaheim, CA
8	Polk	Des Moines-West Des Moines, IA
9	Hennepin	Minneapolis-St. Paul-Bloomington, MN-WI
10	Oklahoma	Oklahoma City, OK



# Geographic PageRank for US CBSAs

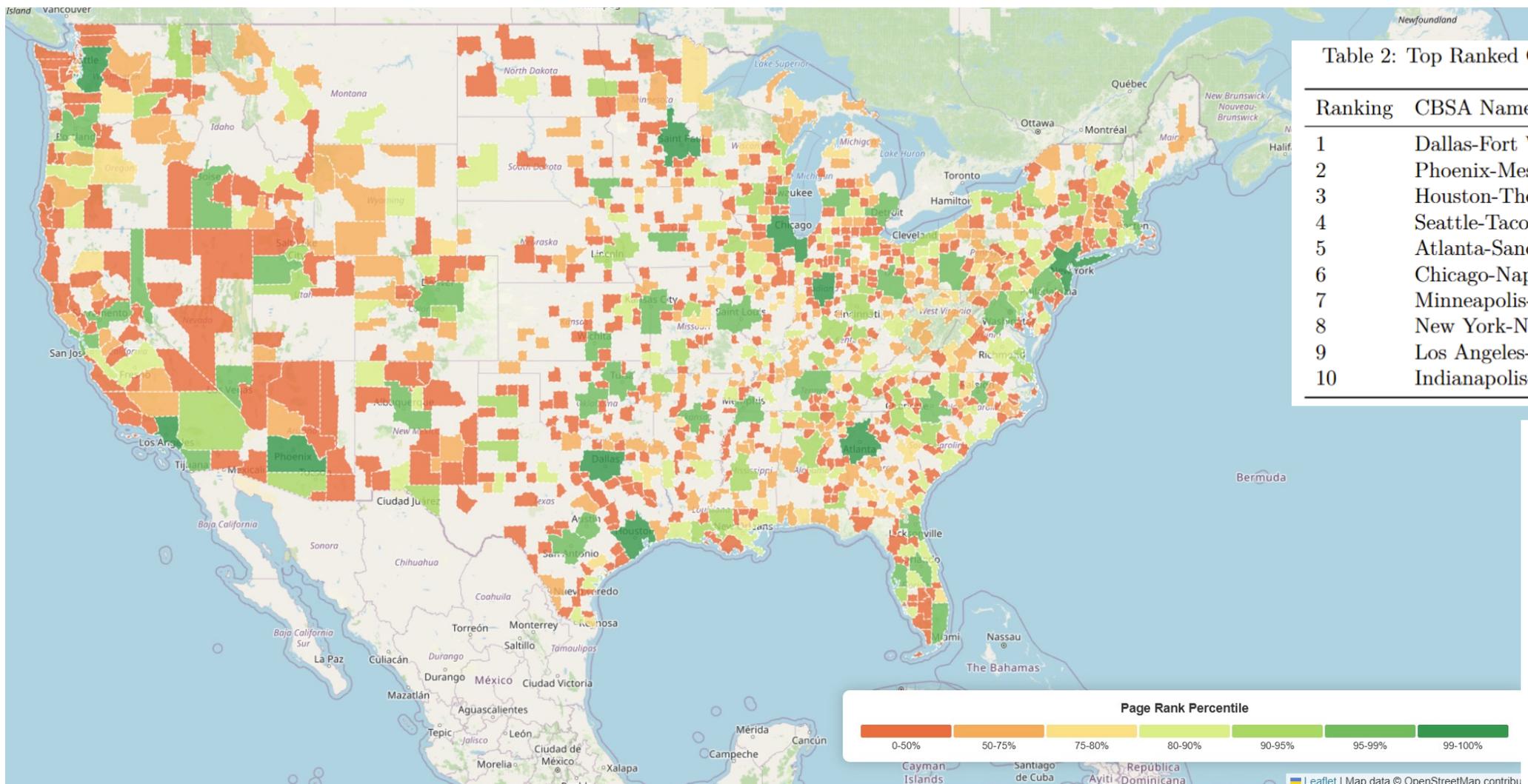


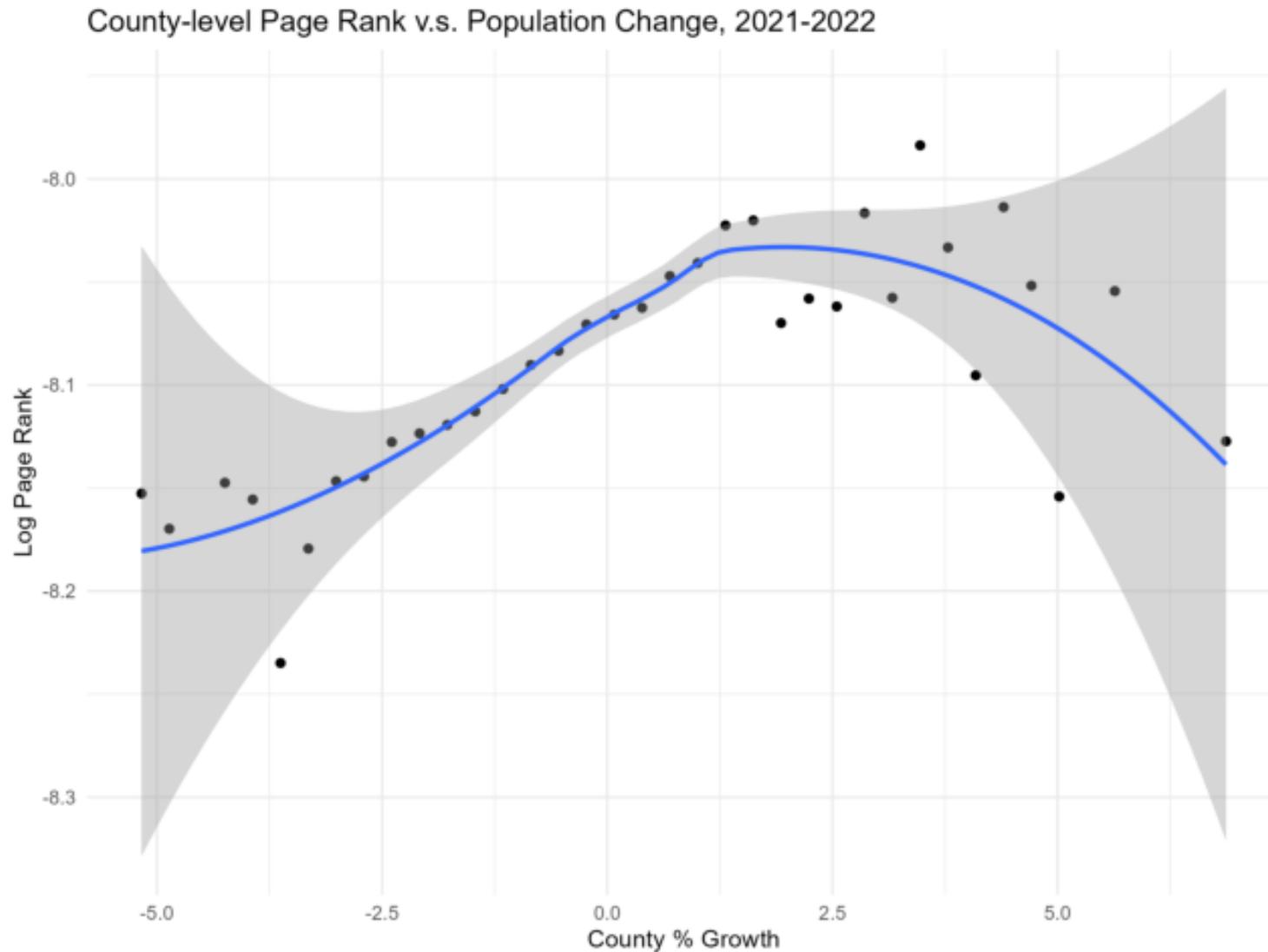
Table 2: Top Ranked CBSAs in the U.S. (2021-2022)

Ranking	CBSA Name
1	Dallas-Fort Worth-Arlington, TX
2	Phoenix-Mesa-Chandler, AZ
3	Houston-The Woodlands-Sugar Land, TX
4	Seattle-Tacoma-Bellevue, WA
5	Atlanta-Sandy Springs-Alpharetta, GA
6	Chicago-Naperville-Elgin, IL-IN-WI
7	Minneapolis-St. Paul-Bloomington, MN-WI
8	New York-Newark-Jersey City, NY-NJ-PA
9	Los Angeles-Long Beach-Anaheim, CA
10	Indianapolis-Carmel-Anderson, IN



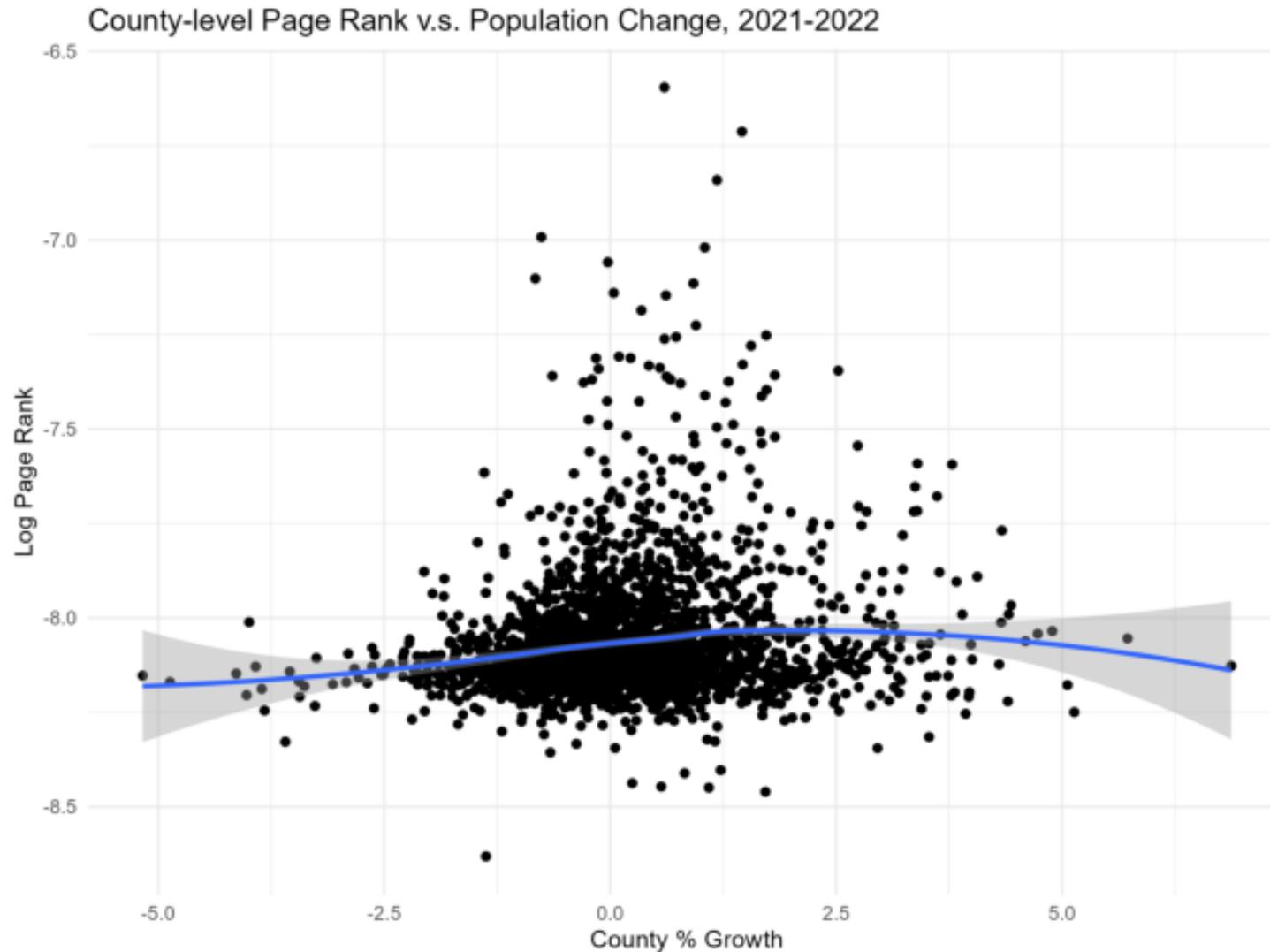
<https://sophieqzwang.github.io/geopagerank/>

# Comparison with Net Migration Rates



- Places with more inflows are ranked higher
- Places with more inflows from other higher-ranked places are ranked even higher

# Comparison with Net Migration Rates



- Places with more inflows are ranked higher
- Places with more inflows from other higher-ranked places are ranked even higher
- Still significant dispersion compared to pure migration measures

# Outline

- I. Motivation
- II. Theory of the PageRank Algorithm
- III. Empirics and Measurement
  - a) Cross-sectional rankings
  - b) Time-series rankings
  - c) Rankings by demographic subgroups
- IV. Application for pricing amenities

# Geographic PageRank for US Counties: 2001 - 2022

Top 30 Metros in 2001–2002: Rank Change by 2021–2022

Metro	Rank 2001–2002	Rank 2021–2022	Improvement
Charlotte-Concord-Gastonia, NC-SC	25	17	+8
Nashville-Davidson–Murfreesboro–Franklin, TN	27	19	+8
Raleigh-Cary, NC	30	23	+7
New York-Newark-Jersey City, NY-NJ	14	8	+6
San Antonio-New Braunfels, TX	19	13	+6
Boston-Cambridge-Newton, MA-NH	21	16	+5
Atlanta-Sandy Springs-Roswell, GA	8	5	+3
Oklahoma City, OK	17	14	+3
Columbus, OH	13	11	+2
Tampa-St. Petersburg-Clearwater, FL	23	21	+2
Seattle-Tacoma-Bellevue, WA	5	4	+1
Indianapolis-Carmel-Greenwood, IN	11	10	+1
Dallas-Fort Worth-Arlington, TX	1	1	0
Phoenix-Mesa-Chandler, AZ	2	2	0
Houston-Pasadena-The Woodlands, TX	3	3	0
Chicago-Naperville-Elgin, IL-IN	6	6	0
Miami-Fort Lauderdale-West Palm Beach, FL	20	20	0
Detroit-Warren-Dearborn, MI	22	22	0
Boise City, ID	28	28	0
Los Angeles-Long Beach-Anaheim, CA	7	9	-2
Minneapolis-St. Paul-Bloomington, MN-WI	4	7	-3
Portland-Vancouver-Hillsboro, OR-WA	9	12	-3
Kansas City, MO-KS	12	15	-3
Washington-Arlington-Alexandria, DC-VA-MD-WV	15	18	-3
San Francisco-Oakland-Fremont, CA	29	33	-4
Las Vegas-Henderson-North Las Vegas, NV	18	24	-6
St. Louis, MO-IL	24	35	-11
San Diego-Chula Vista-Carlsbad, CA	16	30	-14
Sacramento-Roseville-Folsom, CA	26	40	-14
Denver-Aurora-Centennial, CO	10	34	-24

- Among top 30 metros, two decades later
  - Ranking of the top places are similar
  - Dallas, Phoenix, and Houston stay as the top 3
- Top risers:
  - Charlotte (25->17)
  - Nashville (27->19)
  - Raleigh (30->23)
- Top fallers:
  - Denver (10->34)
  - Sacramento (26->40)
  - San Diego (16->30)

# Outline

- I. Motivation
- II. Theory of the PageRank Algorithm
- III. Empirics and Measurement
  - a) Cross-sectional rankings
  - b) Time-series rankings
  - c) Rankings by demographic subgroups
- IV. Application for pricing amenities

# Geographic PageRank By Subgroups

- **Data: ACS micro data**

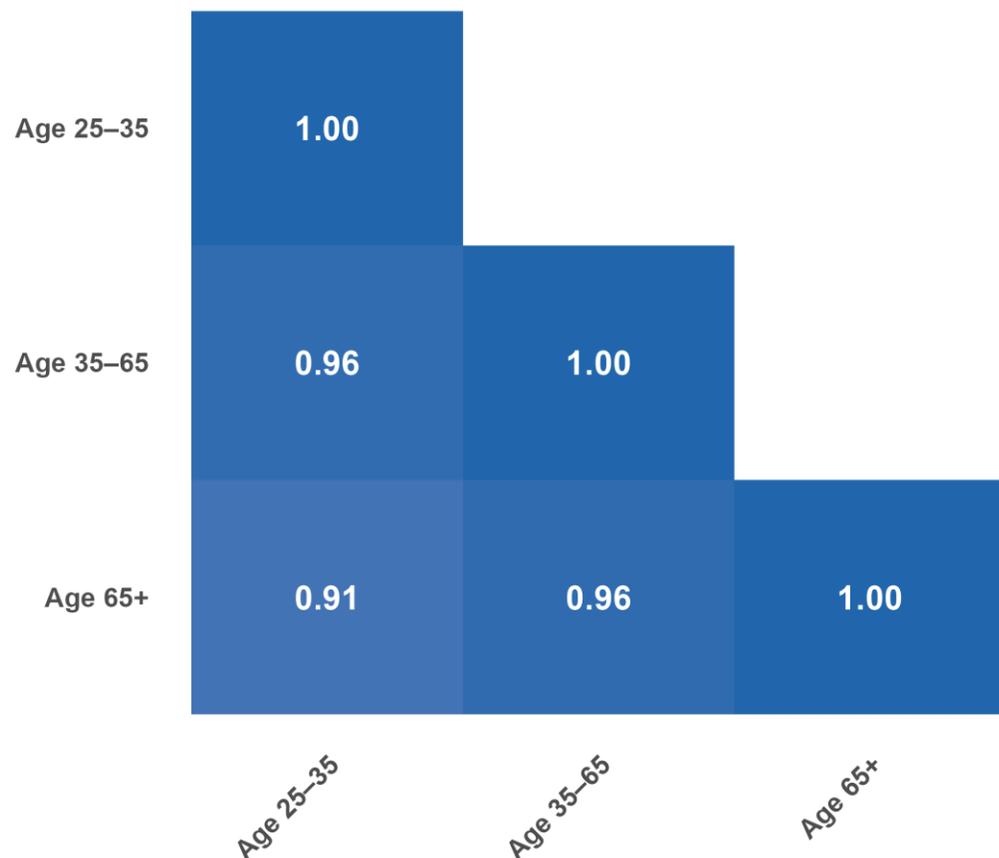
- Includes the migration PUMA from last year (if moved)
  - Aggregate to metro measures to be consistent across years
  - Aggregate for multiple years (5yr) given that ACS is only a sample
- A wide array of household and individual characteristics:
  - Age, Education, Race, and Industry
- The value of a place may vary for these sub-groups

- **Main findings:**

- For most demographic characteristics, the ranking of metros are highly correlated
  - However, important differences remain

# Geographic PageRank for US Metros: By Age Group

- Highly correlated for all age groups
- The rankings for the oldest movers and the youngest movers are least correlated



---

Rank	Age 25-35
1	New York-Newark-Jersey City, NY-NJ-PA
2	Dallas-Fort Worth-Arlington, TX
3	Chicago-Naperville-Elgin, IL-IN-WI
4	Houston-The Woodlands-Sugar Land, TX
5	Los Angeles-Long Beach-Anaheim, CA

---

Rank	Age 35-65
1	Dallas-Fort Worth-Arlington, TX
2	Atlanta-Sandy Springs-Alpharetta, GA
3	Houston-The Woodlands-Sugar Land, TX
4	Phoenix-Mesa-Chandler, AZ
5	Washington-Arlington-Alexandria, DC-VA-MD-WV

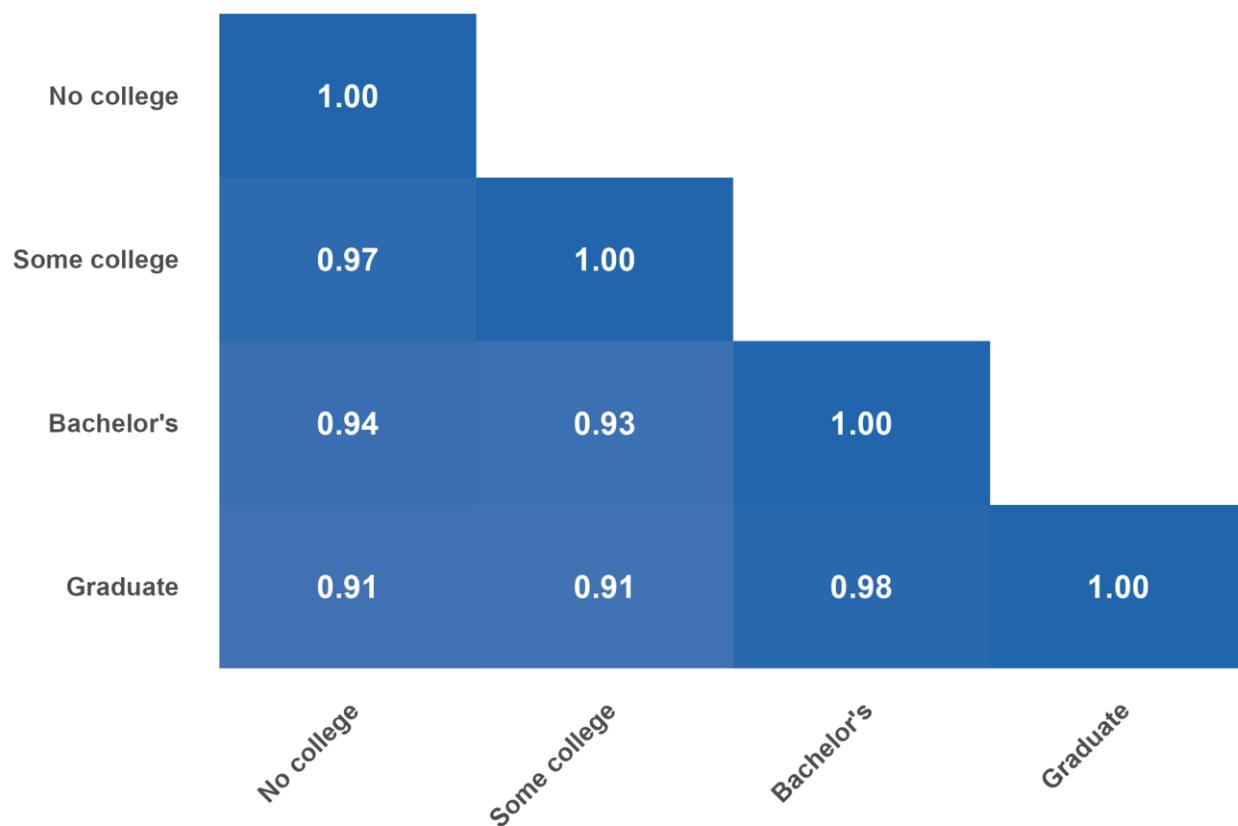
---

Rank	Age 65+
1	Phoenix-Mesa-Chandler, AZ
2	Dallas-Fort Worth-Arlington, TX
3	Atlanta-Sandy Springs-Alpharetta, GA
4	Houston-The Woodlands-Sugar Land, TX
5	Tampa-St. Petersburg-Clearwater, FL

---

# Geographic PageRank for US Metros: By Education

- Highly correlated for levels of education
  - The rankings for those without college are least correlated with those with graduate degrees



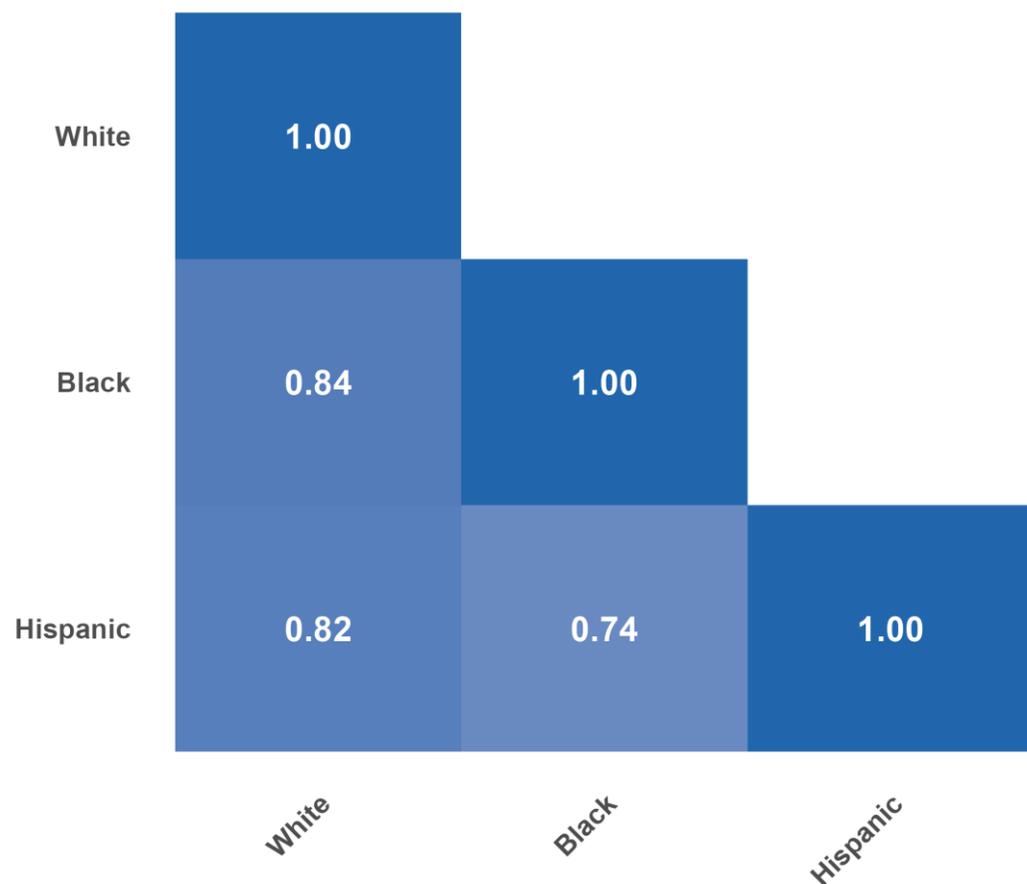
Rank	No College
1	Dallas-Fort Worth-Arlington, TX
2	Houston-The Woodlands-Sugar Land, TX
3	Atlanta-Sandy Springs-Alpharetta, GA
4	Phoenix-Mesa-Chandler, AZ
5	Chicago-Naperville-Elgin, IL-IN-WI

Rank	Bachelor's
1	New York-Newark-Jersey City, NY-NJ-PA
2	Dallas-Fort Worth-Arlington, TX
3	Atlanta-Sandy Springs-Alpharetta, GA
4	Chicago-Naperville-Elgin, IL-IN-WI
5	Los Angeles-Long Beach-Anaheim, CA

Rank	Graduate
1	Washington-Arlington-Alexandria, DC-VA-MD-WV
2	New York-Newark-Jersey City, NY-NJ-PA
3	Atlanta-Sandy Springs-Alpharetta, GA
4	Dallas-Fort Worth-Arlington, TX
5	Chicago-Naperville-Elgin, IL-IN-WI

# Geographic PageRank for US Metros: By Race and Ethnicity

- Mostly correlated across racial and ethnicity groups
  - The rankings for Blacks and Hispanics differ



Rank	White
1	Dallas-Fort Worth-Arlington, TX
2	New York-Newark-Jersey City, NY-NJ-PA
3	Phoenix-Mesa-Chandler, AZ
4	Chicago-Naperville-Elgin, IL-IN-WI
5	Boston-Cambridge-Newton, MA-NH

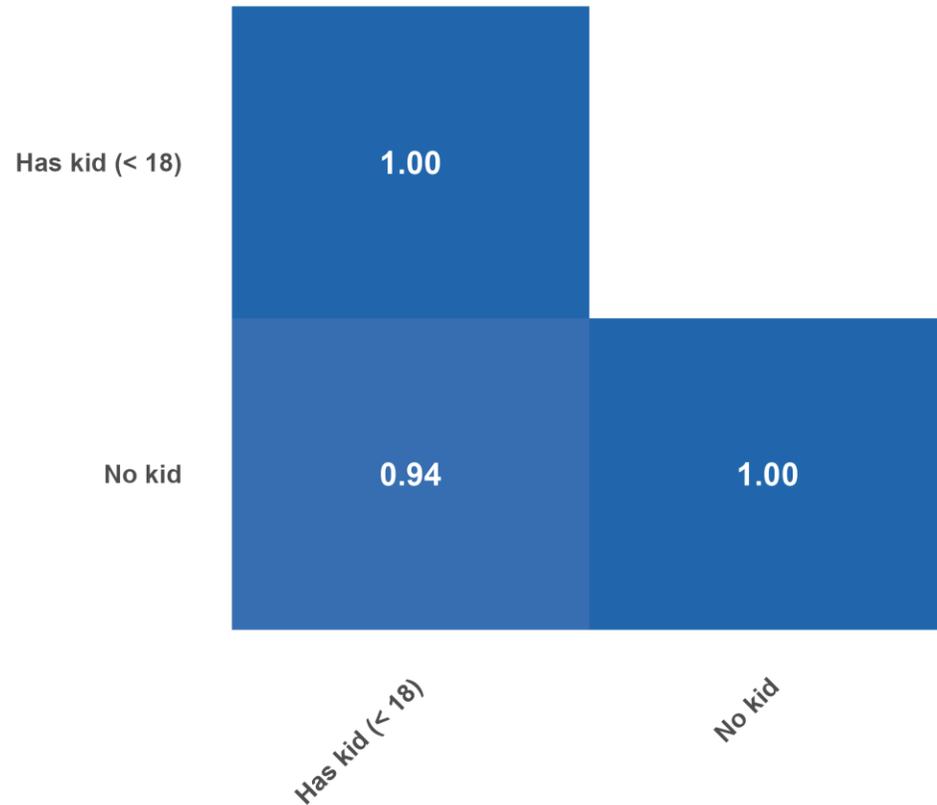
  

Rank	Black
1	Atlanta-Sandy Springs-Alpharetta, GA
2	Houston-The Woodlands-Sugar Land, TX
3	Dallas-Fort Worth-Arlington, TX
4	Washington-Arlington-Alexandria, DC-VA-MD-WV
5	Chicago-Naperville-Elgin, IL-IN-WI

Rank	Hispanic
1	Houston-The Woodlands-Sugar Land, TX
2	Dallas-Fort Worth-Arlington, TX
3	San Antonio-New Braunfels, TX
4	Phoenix-Mesa-Chandler, AZ
5	New York-Newark-Jersey City, NY-NJ-PA

# Geographic PageRank for US Metros: By Family Structure

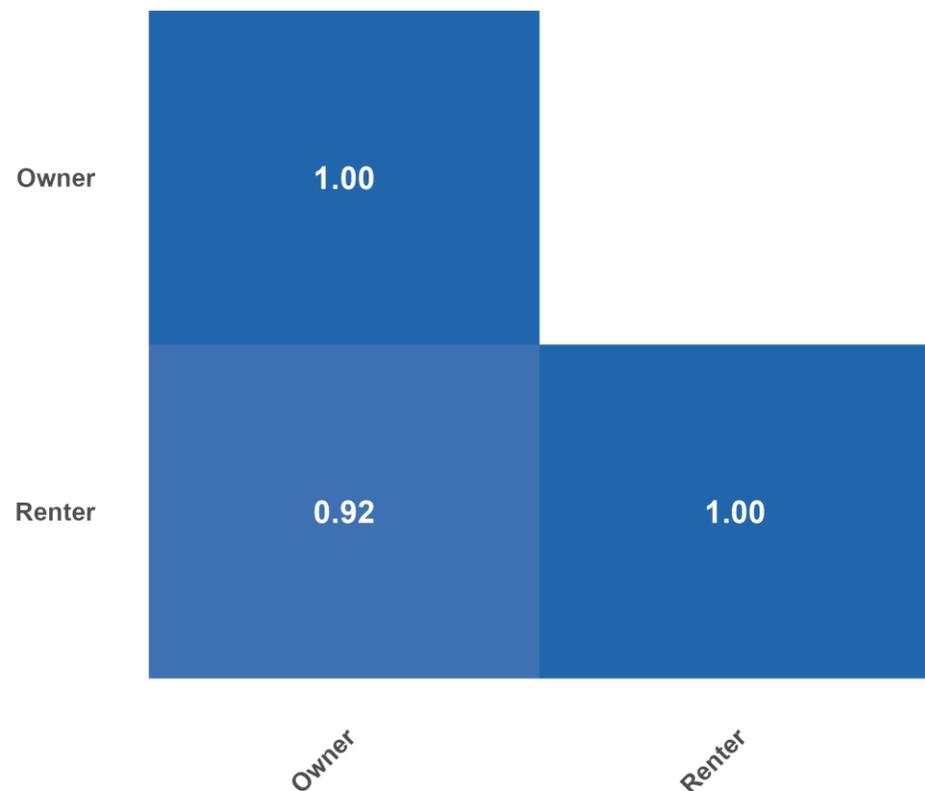


Rank	Has Kid(s) (< 18)
1	Houston-The Woodlands-Sugar Land, TX
2	Dallas-Fort Worth-Arlington, TX
3	Atlanta-Sandy Springs-Alpharetta, GA
4	Chicago-Naperville-Elgin, IL-IN-WI
5	Phoenix-Mesa-Chandler, AZ

Rank	No Kid
1	New York-Newark-Jersey City, NY-NJ-PA
2	Dallas-Fort Worth-Arlington, TX
3	Seattle-Tacoma-Bellevue, WA
4	Los Angeles-Long Beach-Anaheim, CA
5	Houston-The Woodlands-Sugar Land, TX

# Geographic PageRank for US Metros: By Housing Tenure

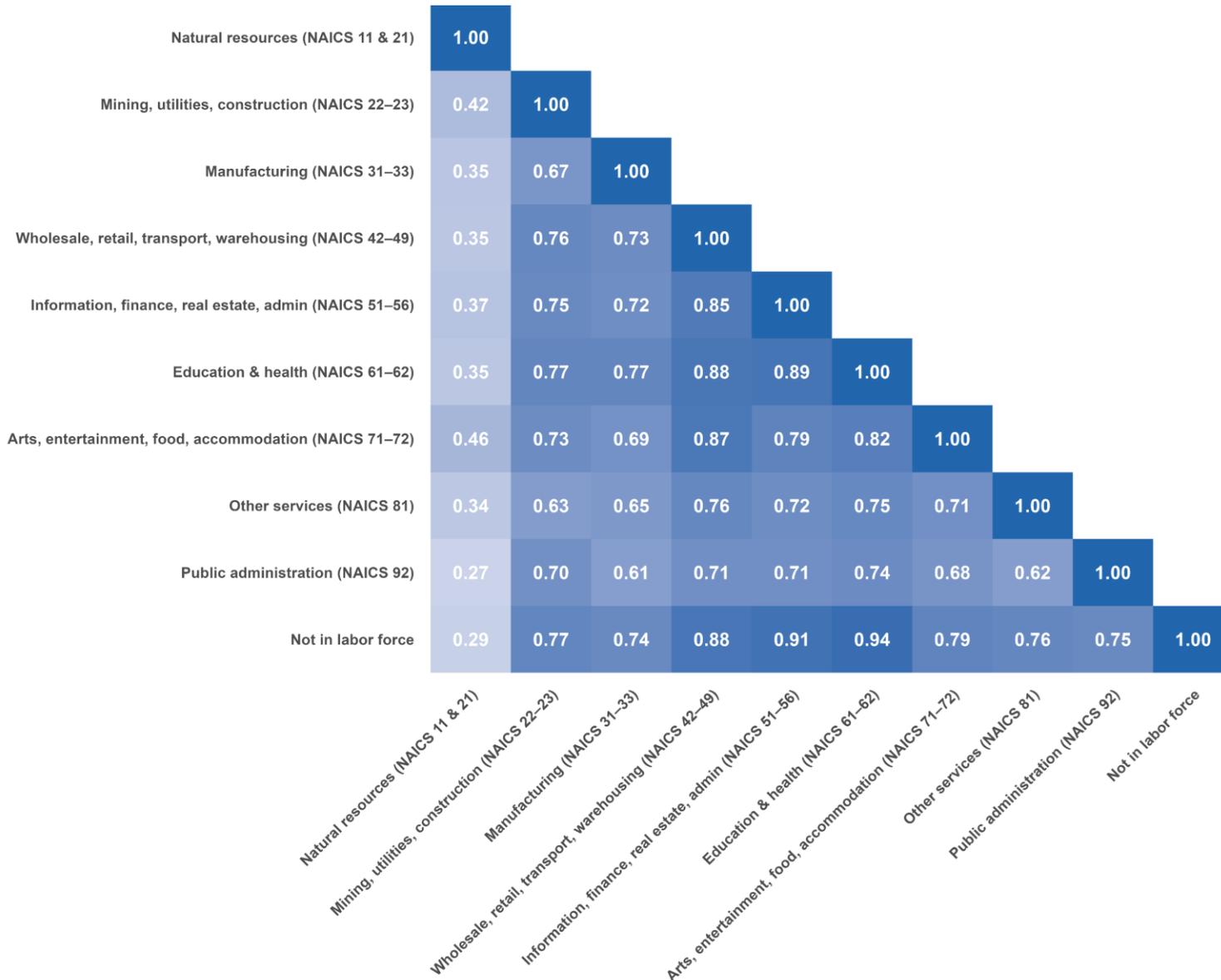


Rank	Owner
1	Atlanta-Sandy Springs-Alpharetta, GA
2	Dallas-Fort Worth-Arlington, TX
3	Houston-The Woodlands-Sugar Land, TX
4	Phoenix-Mesa-Chandler, AZ
5	Chicago-Naperville-Elgin, IL-IN-WI

Rank	Renter
1	Dallas-Fort Worth-Arlington, TX
2	Los Angeles-Long Beach-Anaheim, CA
3	Houston-The Woodlands-Sugar Land, TX
4	New York-Newark-Jersey City, NY-NJ-PA
5	Atlanta-Sandy Springs-Alpharetta, GA

# Geographic PageRank for US Metros: By Industry



- Varying degrees of correlation across industries
  - Industry measured at the destination from the ACS
  - Rankings become noticeably different for those in natural resources, mining, utilities, and manufacturing

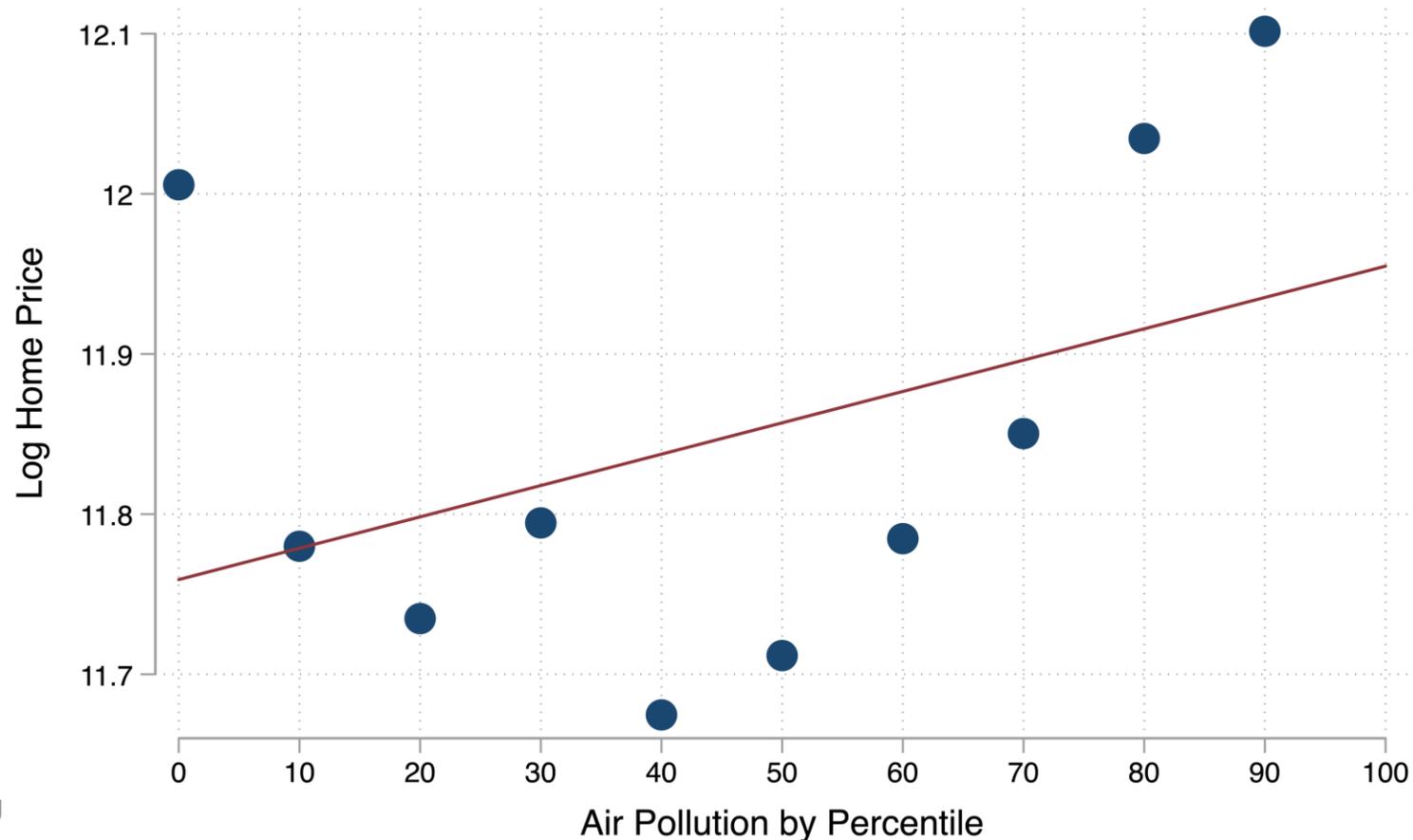
# Outline

- I. Motivation
- II. Theory of the PageRank Algorithm
- III. Empirics and Measurement
  - a) Cross-sectional rankings
  - b) Time-series rankings
  - c) Rankings by demographic subgroups
- IV. Application for pricing amenities

# Application: Pricing Neighborhood Amenities

- **Research question:**

- What is the implicit price of air quality?
  - Empirical challenge: amenities such as clean air can often be correlated with the unobserved quality of the place -> **"The wrong-signed problem"**



# Application: Pricing Neighborhood Amenities

- **Research question:**

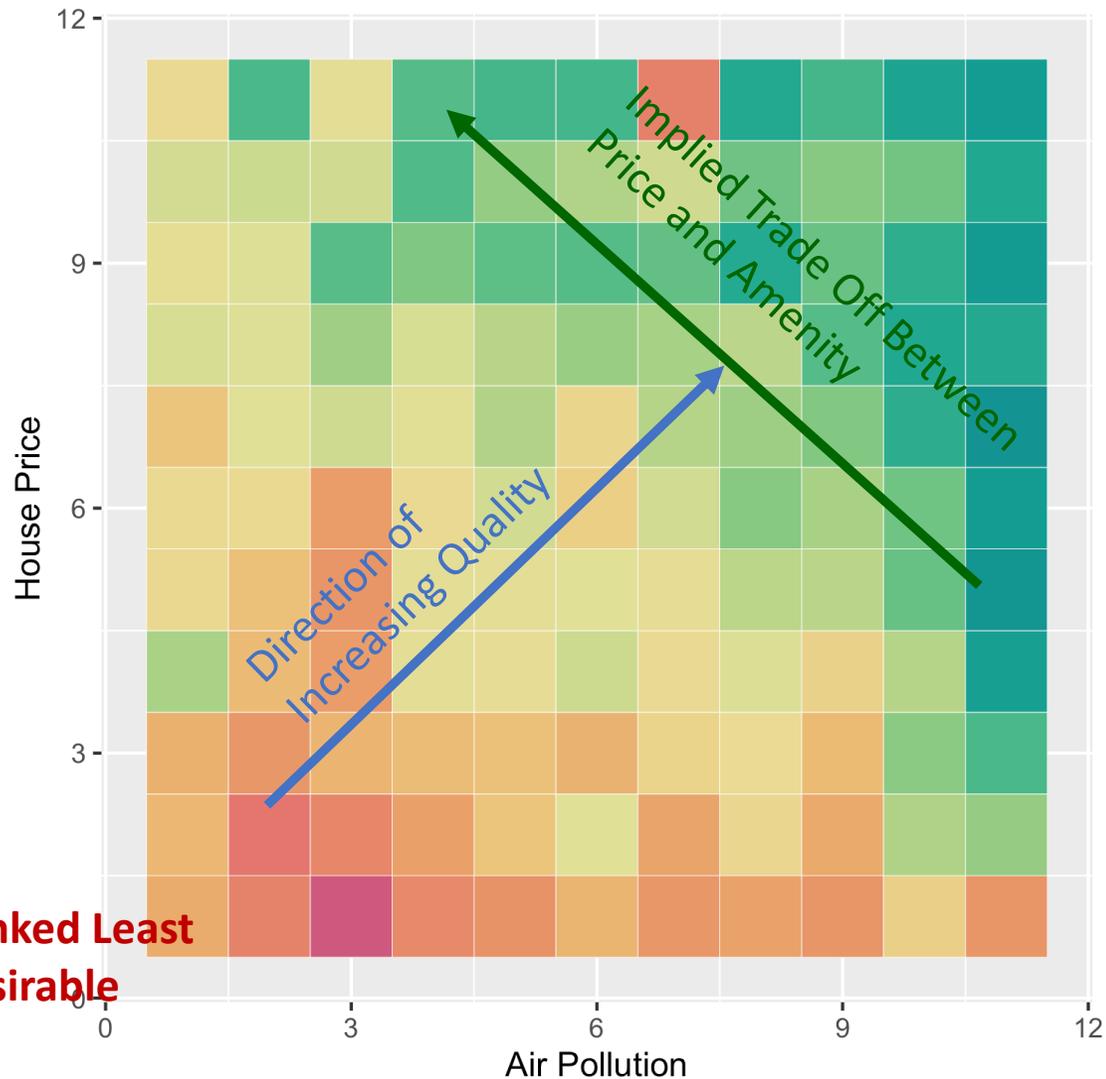
- What is the implicit price of air quality?
  - Empirical challenge: amenities such as clean air can often be correlated with the unobserved quality of the place -> "The wrong-signed problem"

- **Method:**

- Conventional approach:
  - Find more controls -> But can never exhaust all controls
  - Find quasi-experiments -> Limited in scope
- The Anti-IV Approach (Bell, Billings, Calder-Wang, and Zhong, 2024)
  - County-level Geographic PageRank can be used as an anti-instrument
    - GPR co-moves with unobserved housing quality
    - The extent to which GPR mismeasures true housing quality is uncorrelated with air pollution

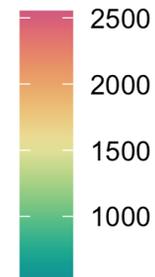
# Application: Pricing Neighborhood Amenities

3D Binscatter at county-level, nationwide 2019



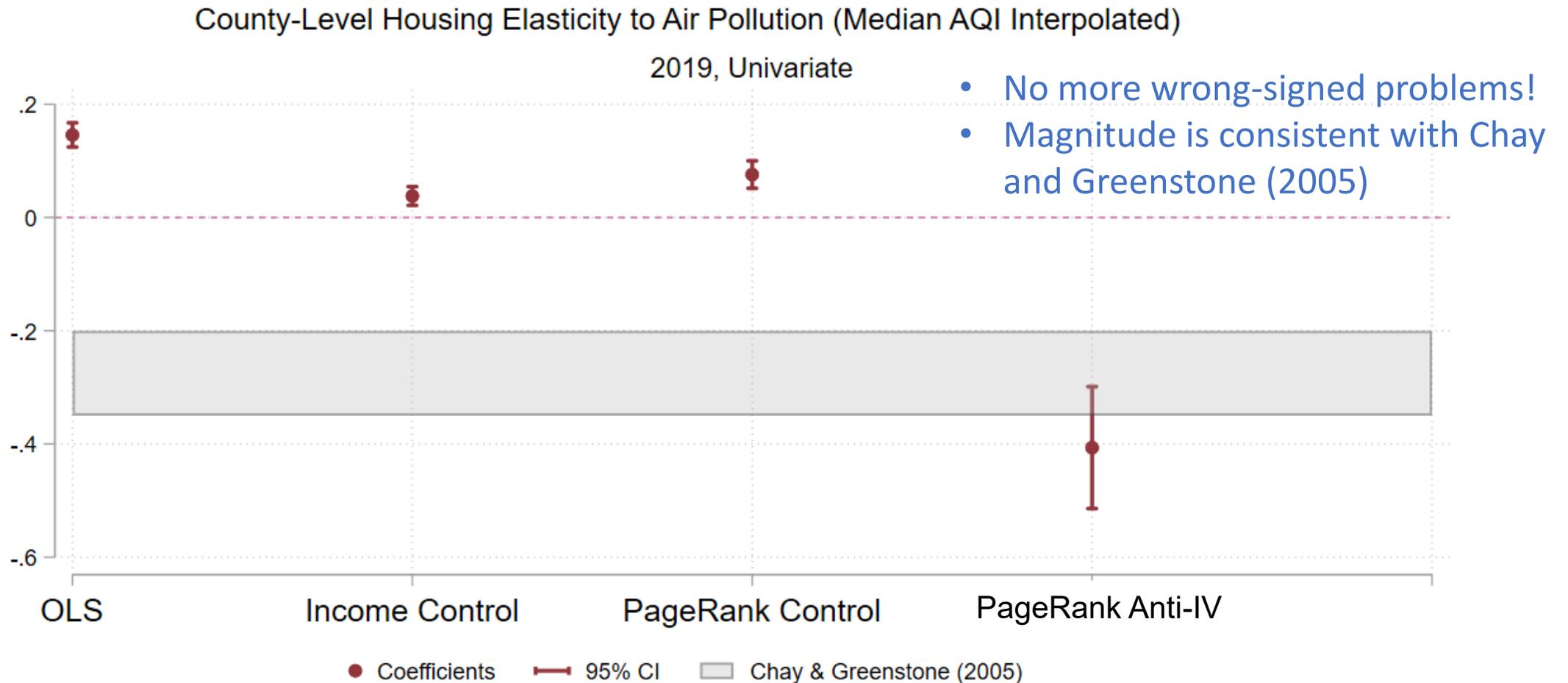
Ranked Most  
Desirable

PageRank Ranking



- X-axis: Deciles of air pollution
- Y-axis: Deciles of home prices
- Z-axis (color): Deciles of predicted Geographic PageRank:
  - $E[ H | P, Z ]$
  - Calculated as the average GPR for the bin

# Application: Pricing Neighborhood Amenities



# Conclusion

- **Theory:**

- Use the well-known PageRank algorithm to rank places
  - A recursive measure based on network centrality
  - Interpreted as the value of place through revealed preference

- **Measurement:**

1. Cross-sectional measures: County-level and metro-level GPR
2. Time-series measures: County-level GPR over time
3. Geographic PageRank for different demographic groups

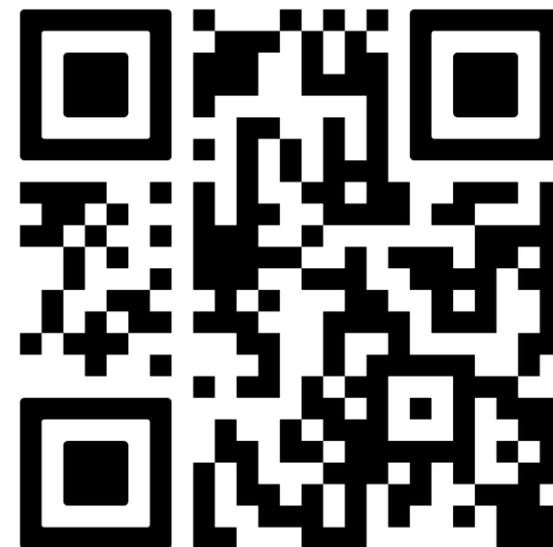
- **Application:**

- Use GPR as an anti-IV for unmeasured quality of housing to price neighborhood amenities
  - Works well for obtaining the implicit price of air quality

Data and Visualization

Available at:

<https://sophieqzwang.github.io/geopagerank/>



# Thanks!

Questions or comments?