



Urban Dreams of Migrants

A Case Study of Migrant Integration in Shanghai

Yang Yang¹, Chenhao Tan², Zongtao Liu¹,
Fei Wu¹, and Yueting Zhuang¹

¹Zhejiang University, ²University of Colorado Boulder

Background

- Migrants make great contribution to the prosperity of cities
- Pose great challenges including **segregation** and **social inequality**
- Understanding migrant integration helps policymakers with **urban planning**

Background

- In China, the fraction of population dwelling in cities increased from 17.9% to 52.6% between 1978 and 2012.
- In China, 260 million people migrate to cities to realize their urban dreams every year.

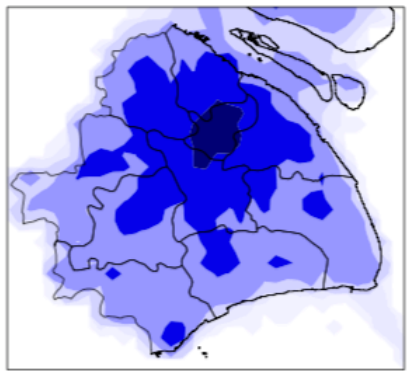
Highlights of Our Work

- Our work presents the **first large-scale quantitative exploration** of migrant integration.
 - *Dataset:*
 - One-month **complete** anonymized telecommunication records in Shanghai from China Telecom*
 - Span 4 weeks from Sep. 3rd, 2016, to Sep. 30th, 2016
 - Contain **698M** call logs between **54M** mobile users, personal attributes of **3M** users, and **GPS** location of each call
- We investigate the evolution of two central components of migrant integration
 - Telecommunication networks
 - Geographical locations

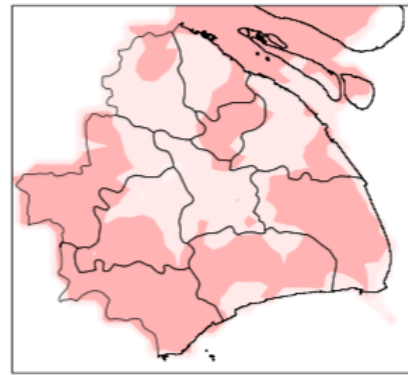
* China Telecom Corporation is a Chinese state-owned telecommunication company and the third largest mobile service providers in China.

Migrants and Locals

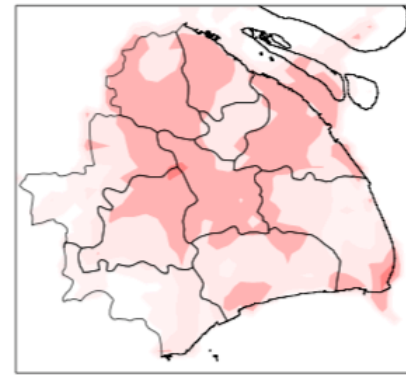
| group | brief | number |
|------------------|---|--------|
| locals | Who were born in Shanghai | 1.7 M |
| settled migrants | Who were not born in Shanghai and had at least one call in the first week | 1.0 M |
| new migrants | Who were not born in Shanghai and had no call in the first week | 22K |



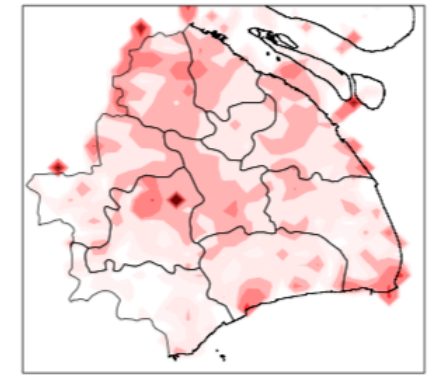
(a) Overall average probability.



(b) Locals.

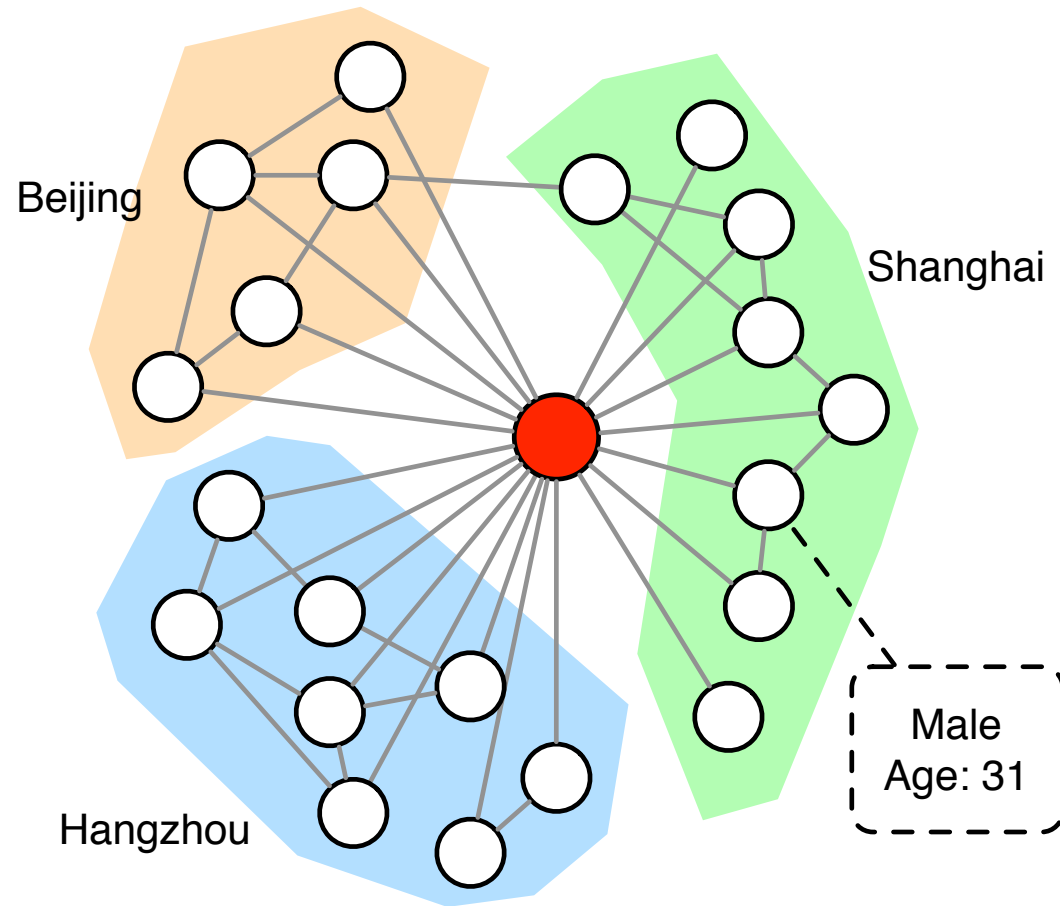


(c) Settled migrants.

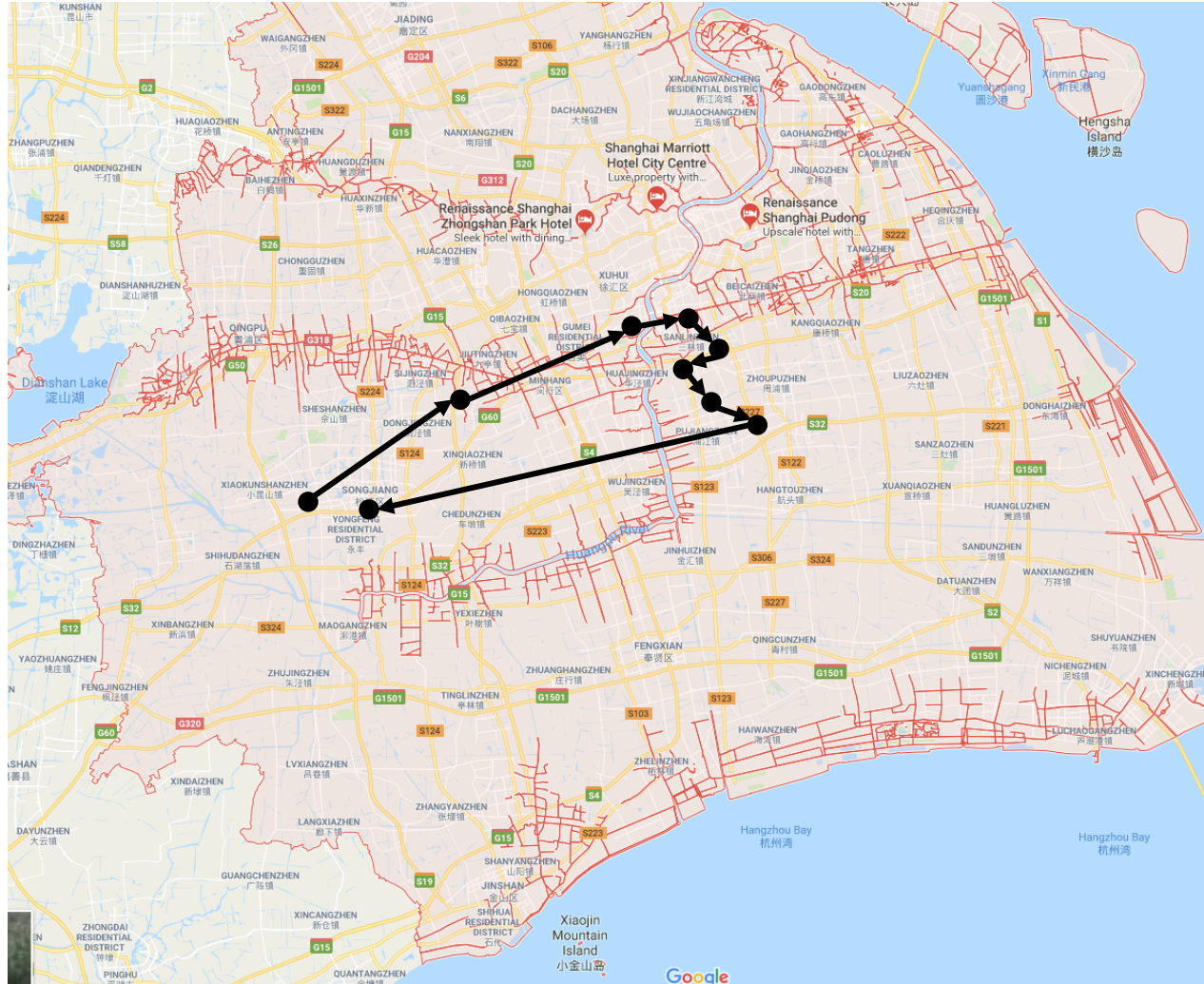


(d) New migrants.

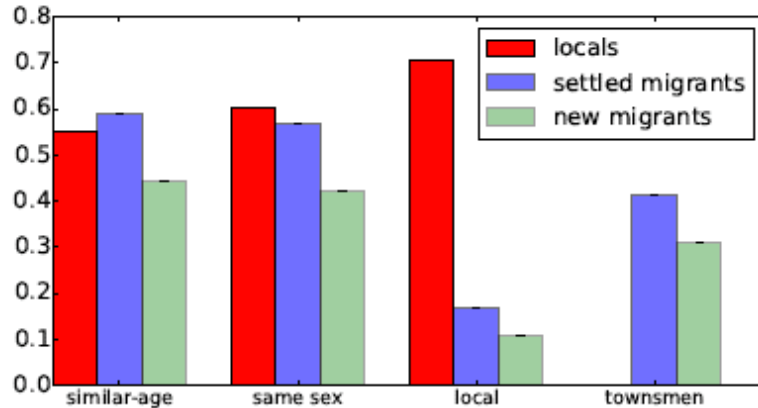
Mobile communication networks



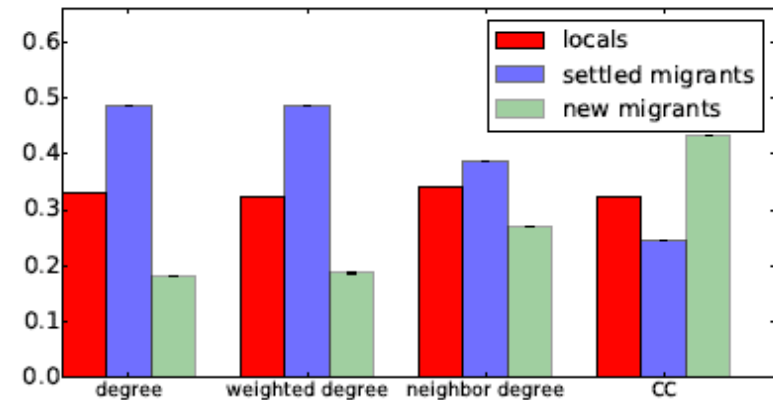
Geographical Locations



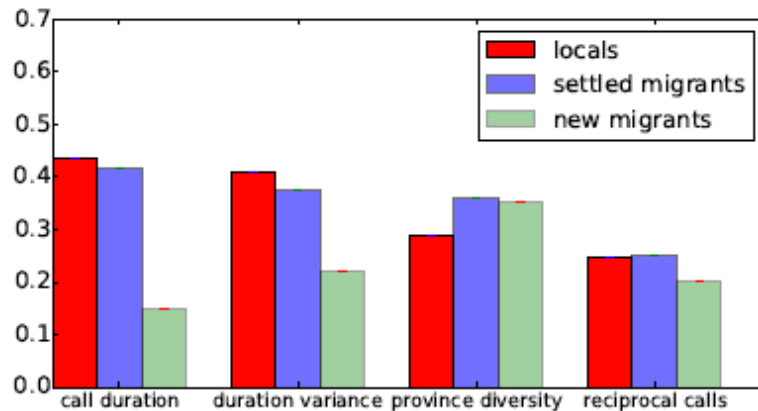
Behavior Patterns in Different Groups



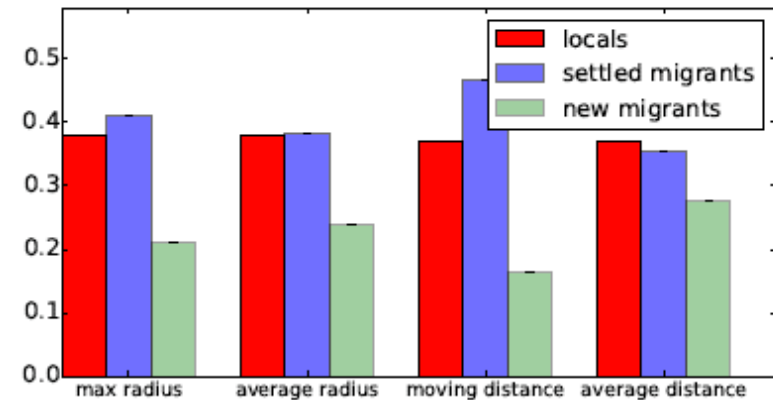
(a) Demographics of friends.



(b) Ego-network characteristics.

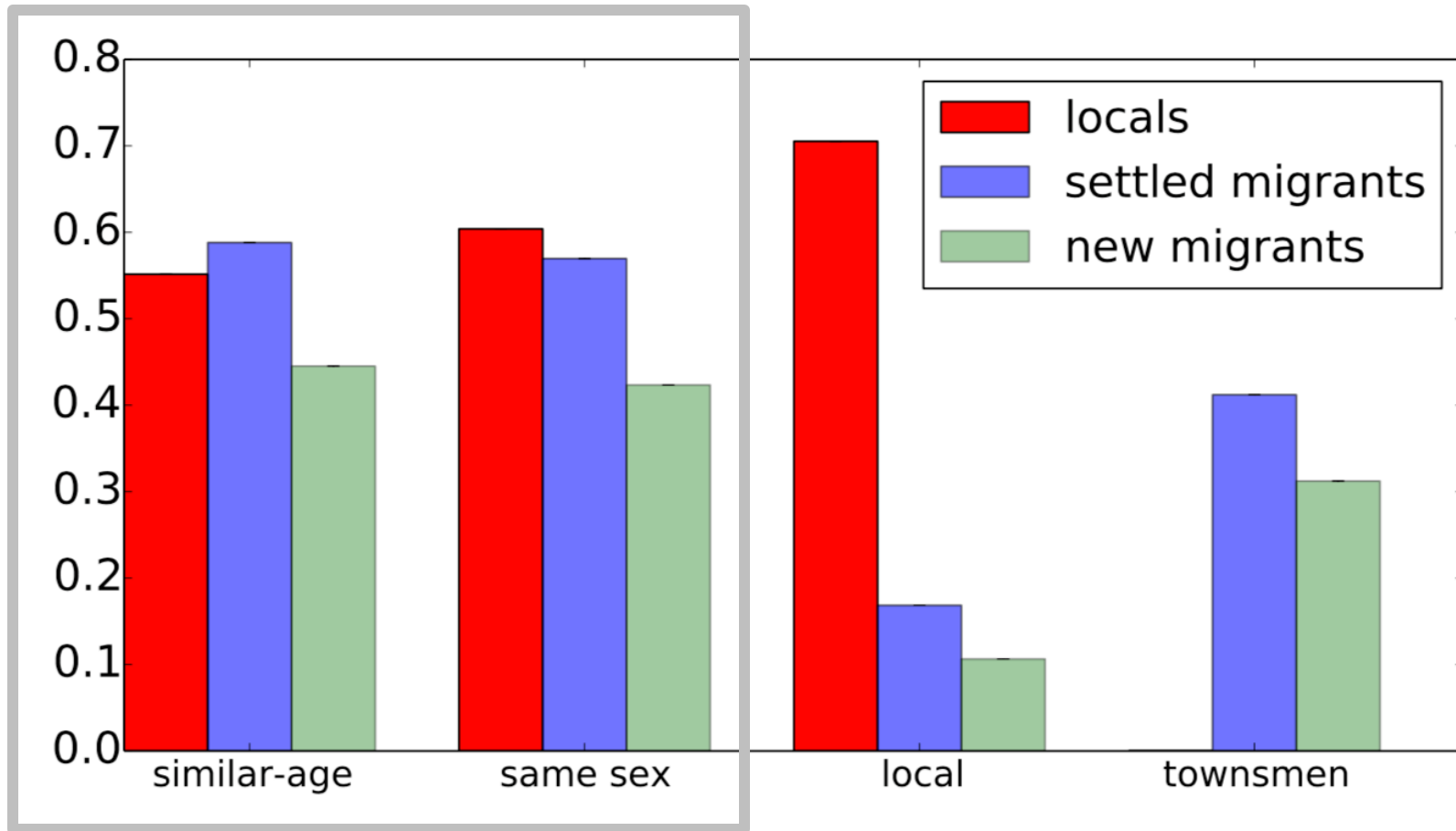


(c) Call behavior.



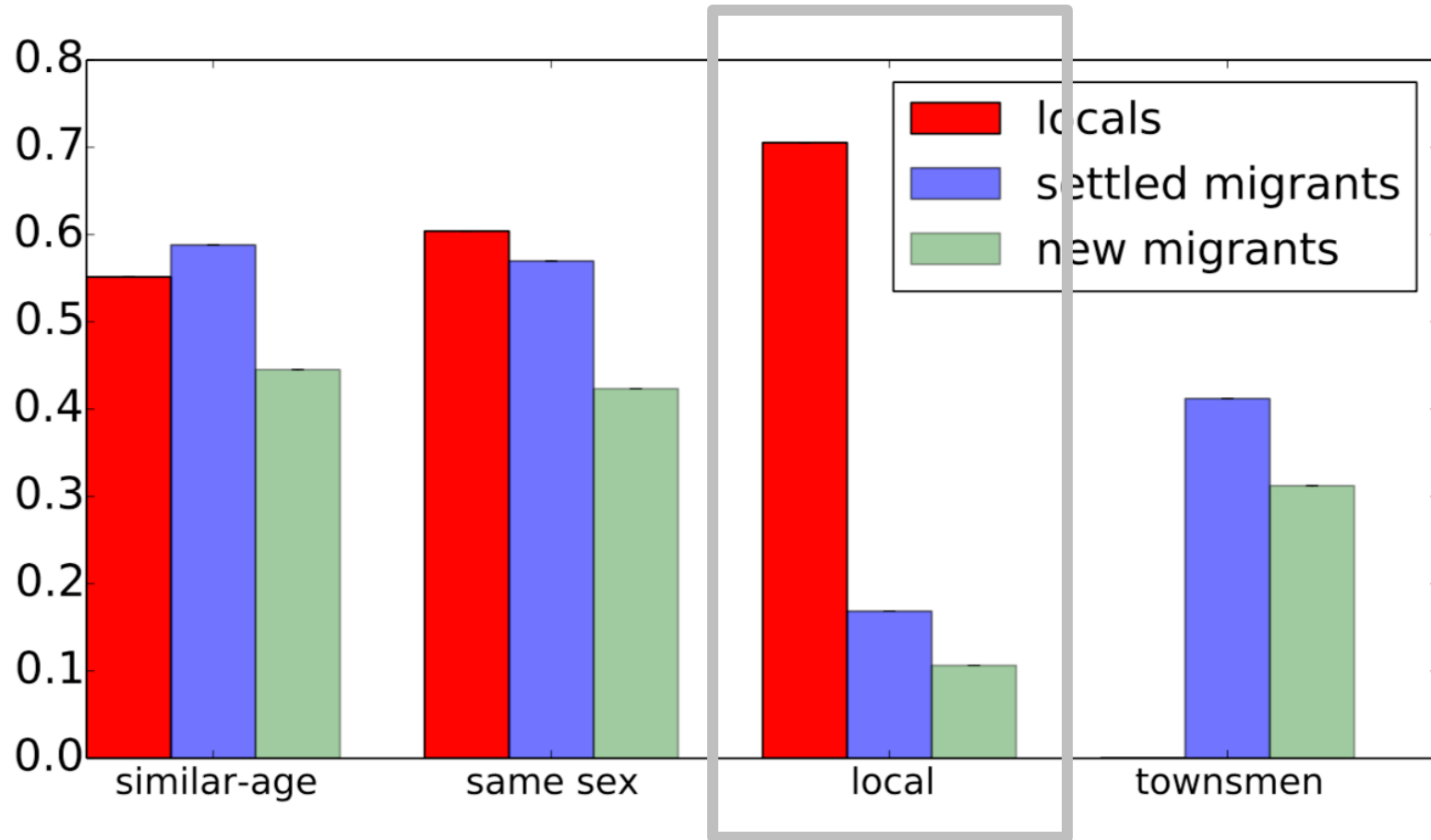
(d) Geographical features.

Behavior Patterns in Different Groups



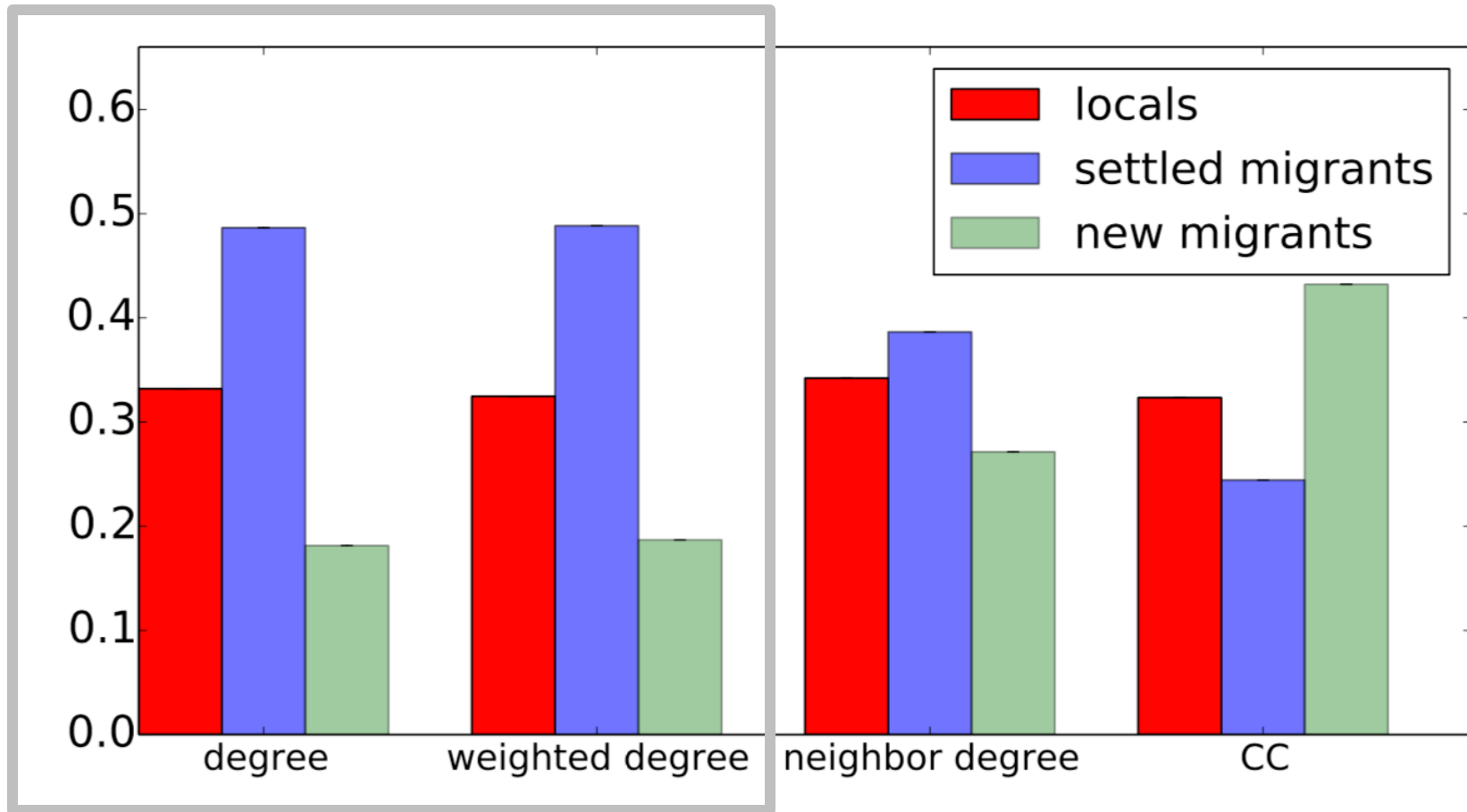
(a) Demographics of friends.

Behavior Patterns in Different Groups



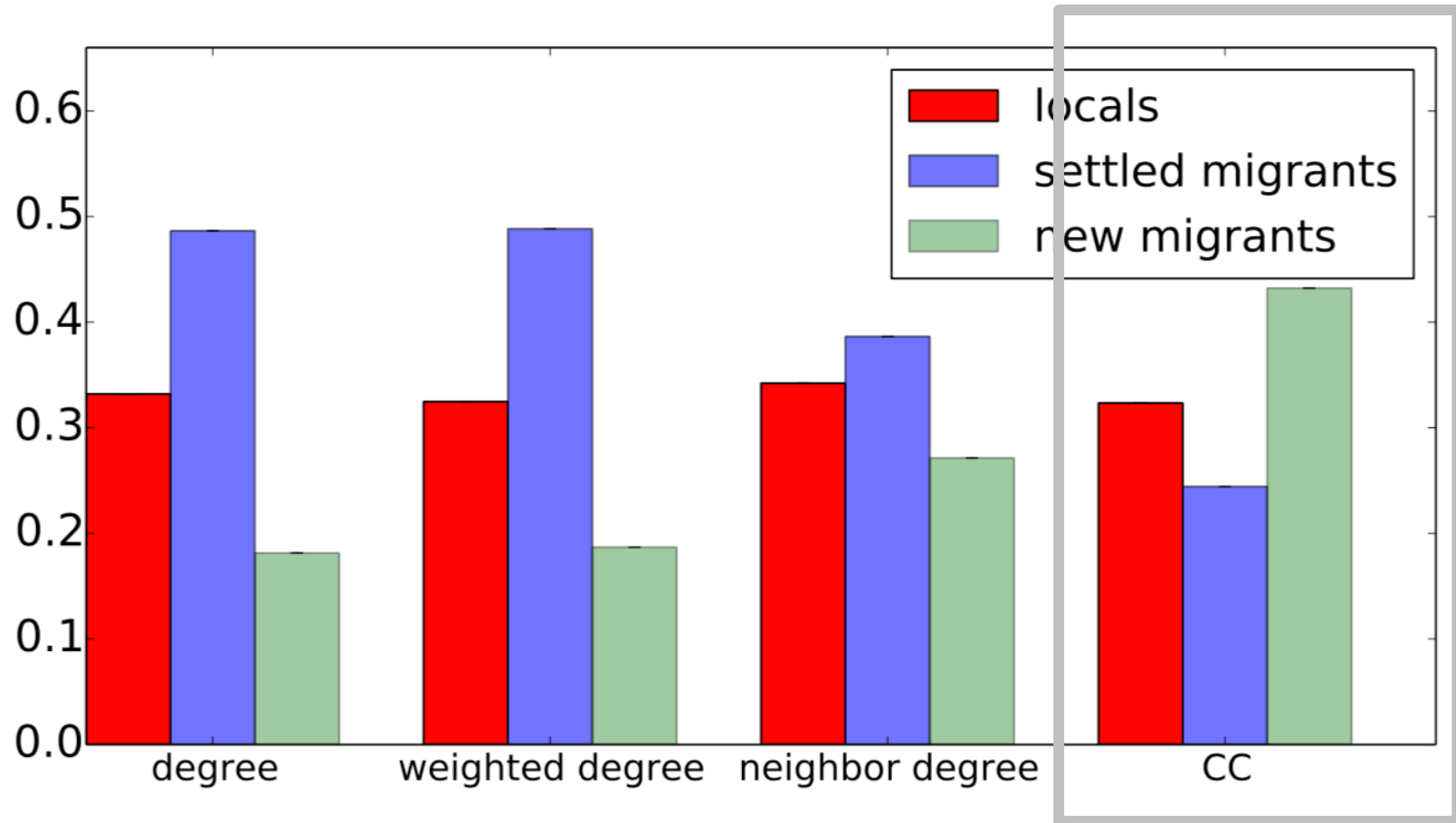
(a) Demographics of friends.

Behavior Patterns in Different Groups



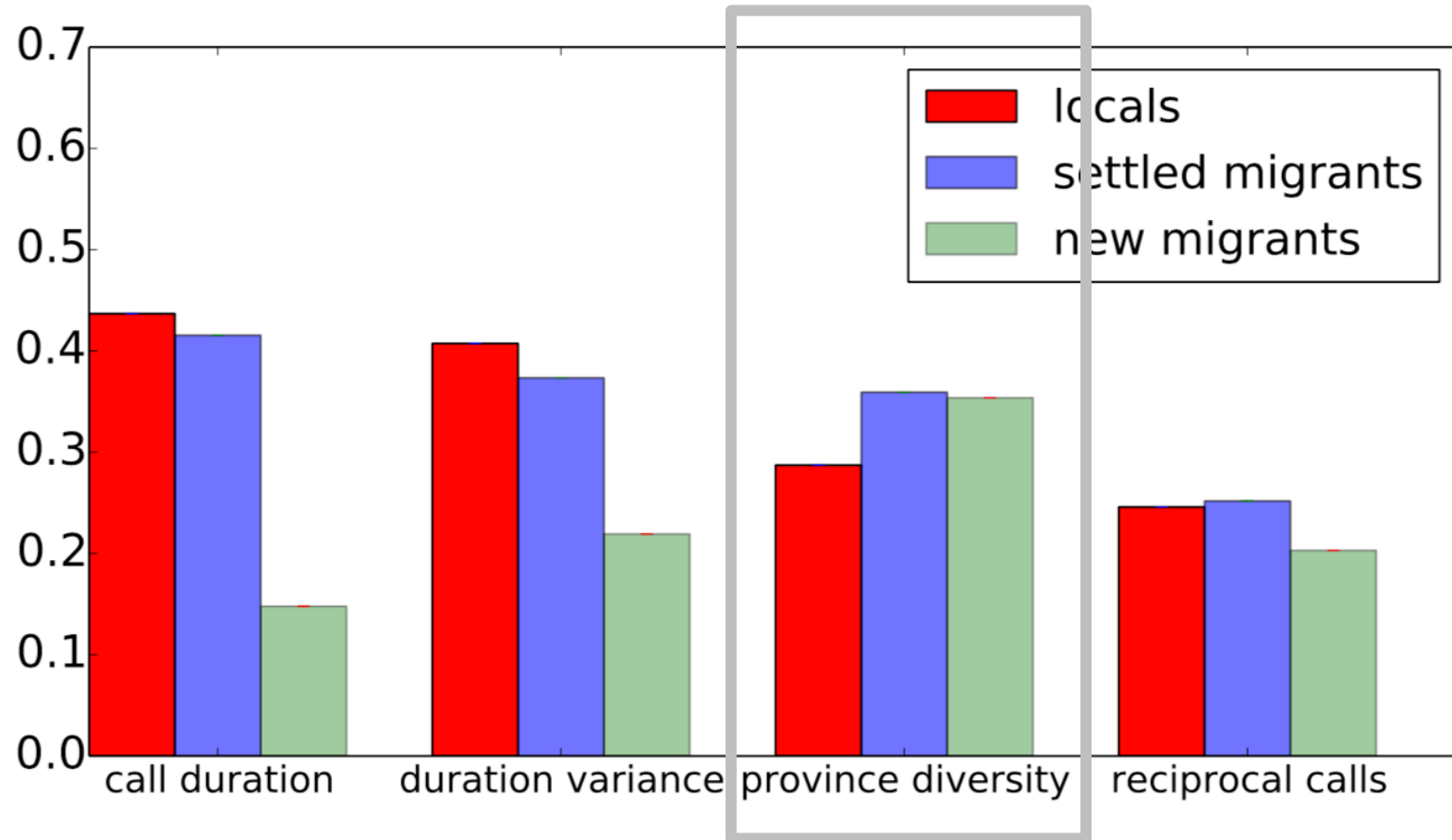
(b) Ego-network characteristics.

Behavior Patterns in Different Groups



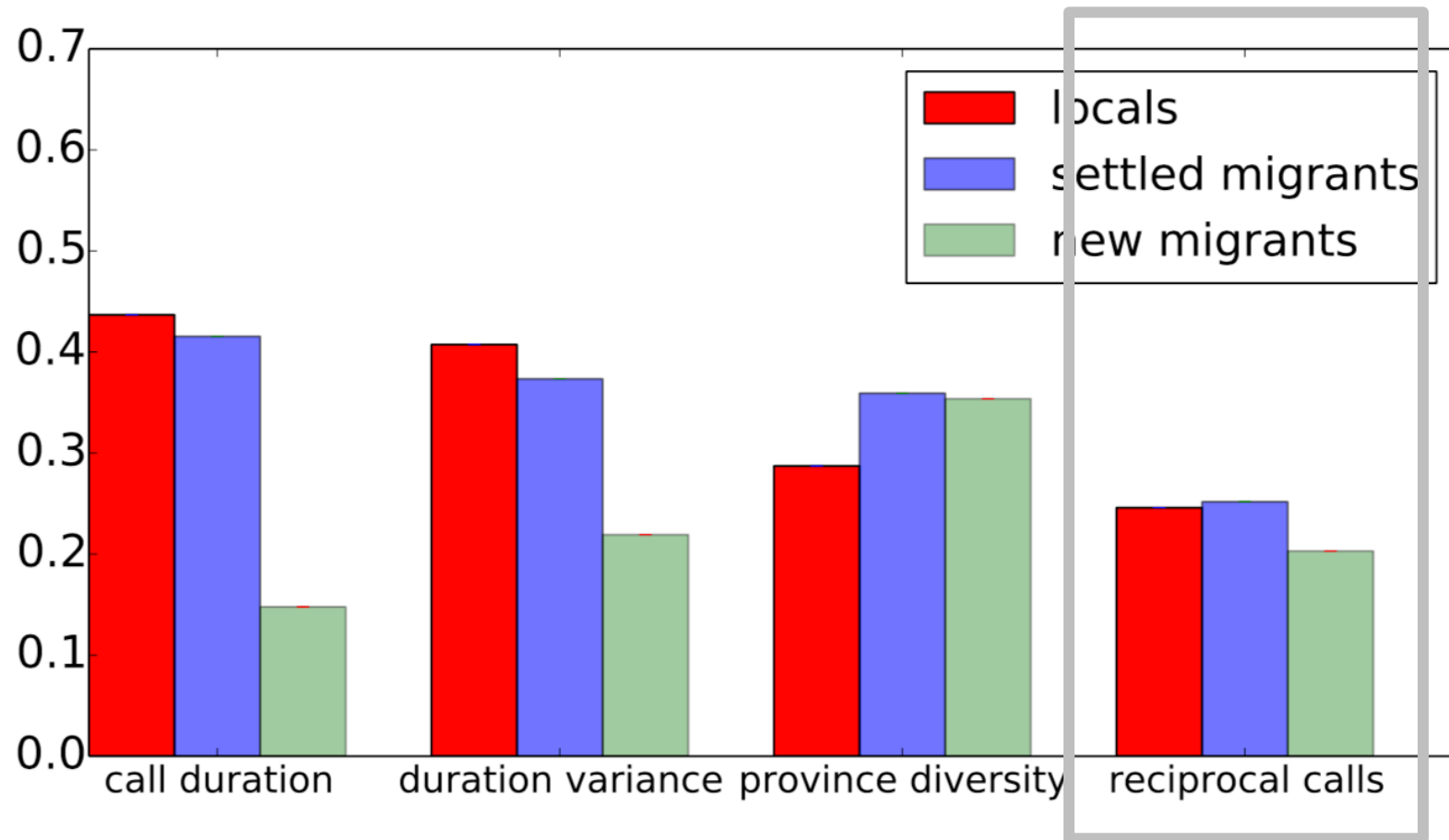
(b) Ego-network characteristics.

Behavior Patterns in Different Groups



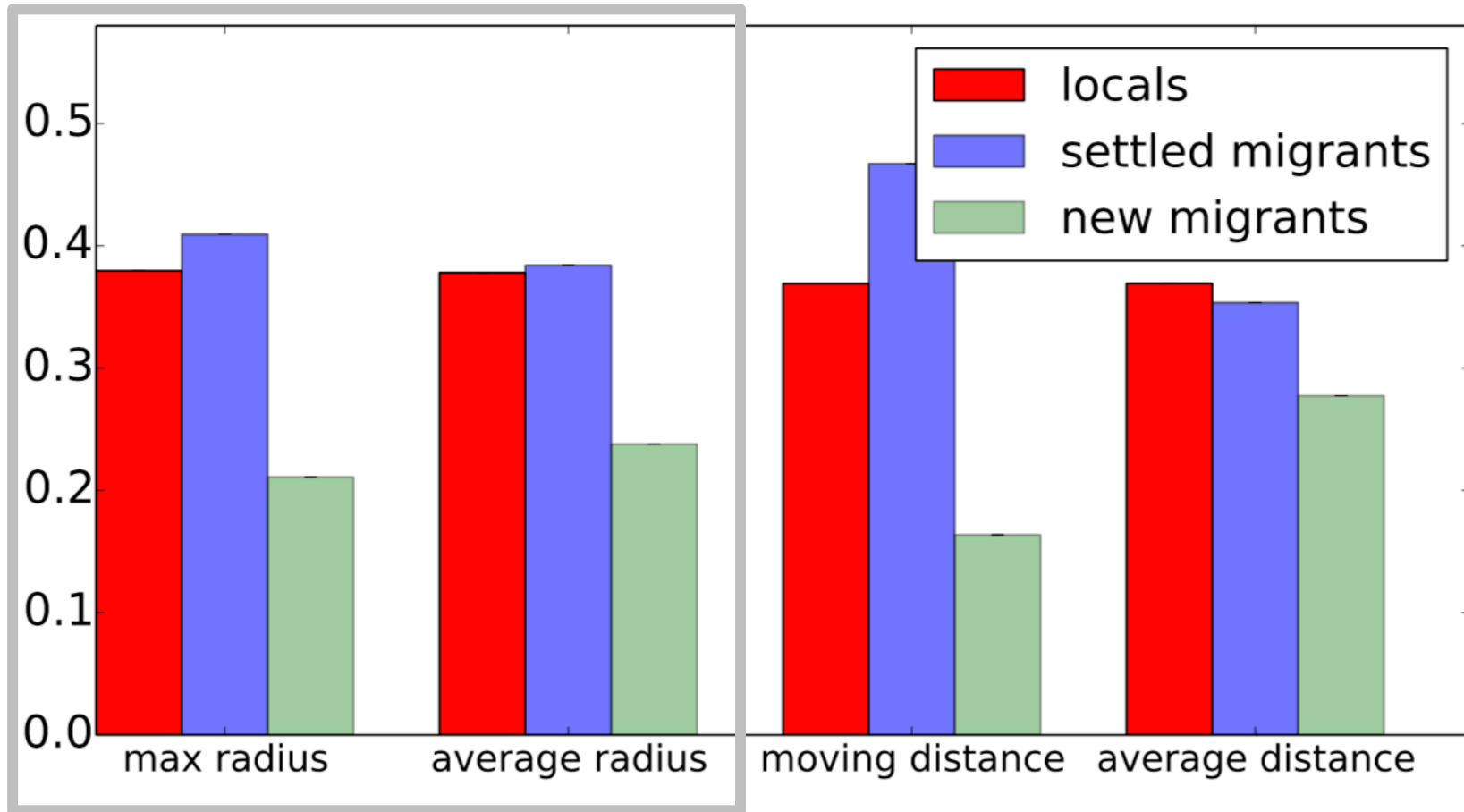
(c) Call behavior.

Behavior Patterns in Different Groups



(c) Call behavior.

Behavior Patterns in Different Groups



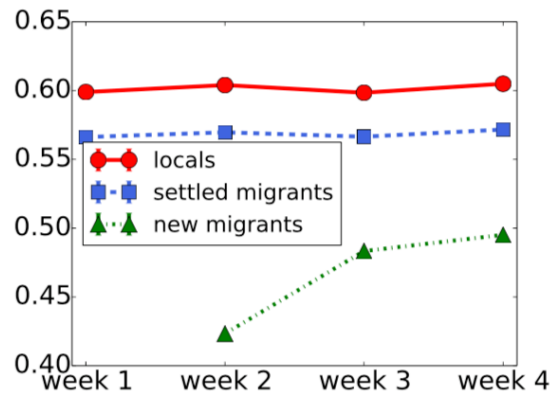
(d) Geographical features.

Integration of New Migrants

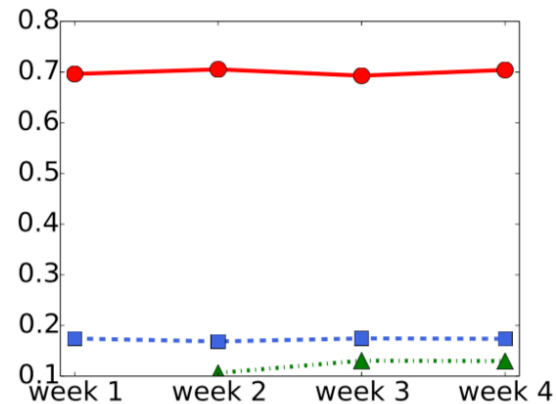
New migrants → Settled migrants → Locals

Integration of New Migrants

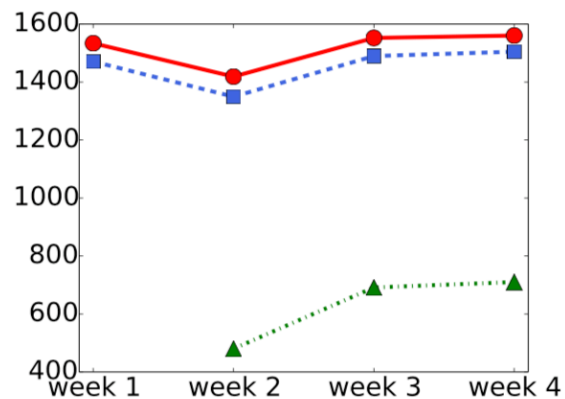
- ① new migrants move towards settled migrants and both move towards locals



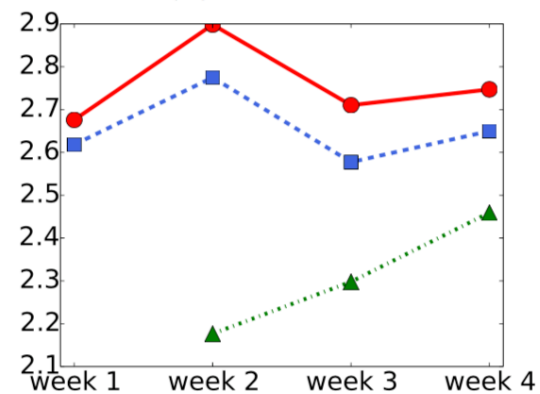
(a) Same-sex.



(b) Local.



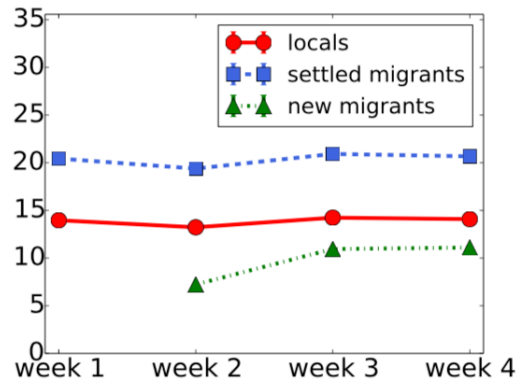
(c) Call duration (in seconds).



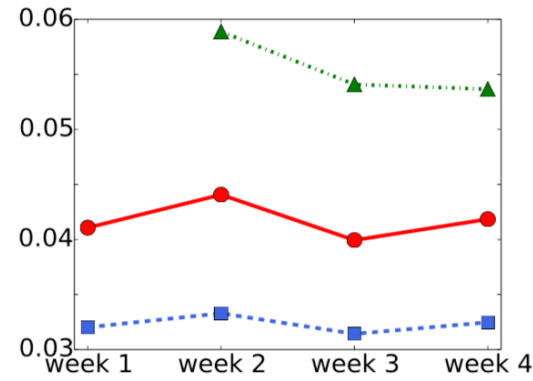
(d) Average moving distance.

Integration of New Migrants

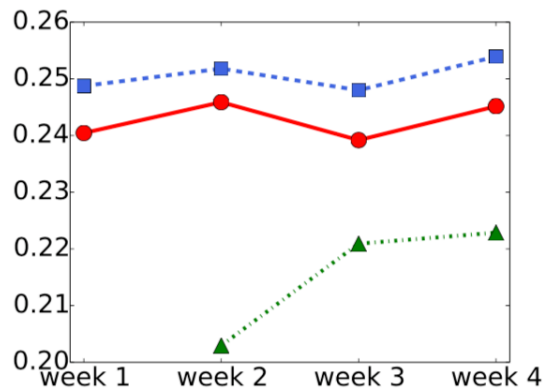
- ② new migrants move towards locals initially, but will likely eventually move away from locals



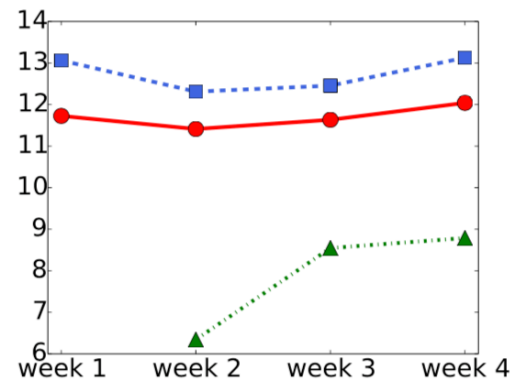
(e) Degree.



(f) Clustering coefficient.



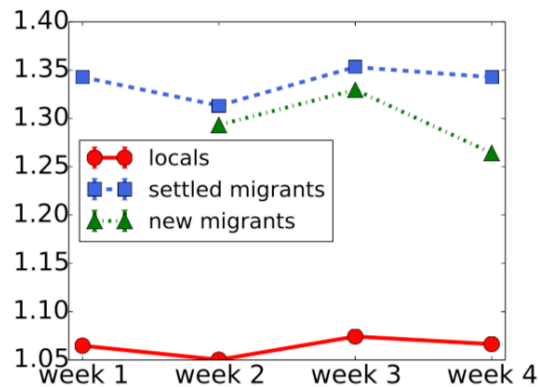
(g) Reciprocal calls.



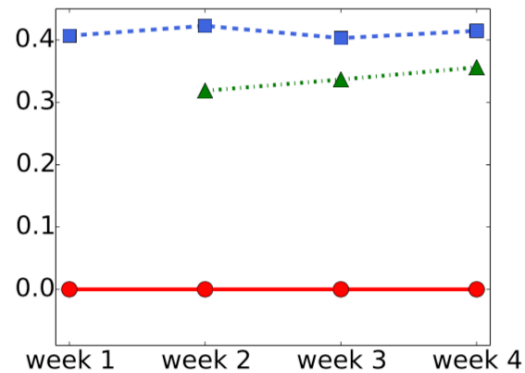
(h) Max radius.

Integration of New Migrants

- ③ new migrants move towards settled migrants but away from locals



(i) Province diversity.



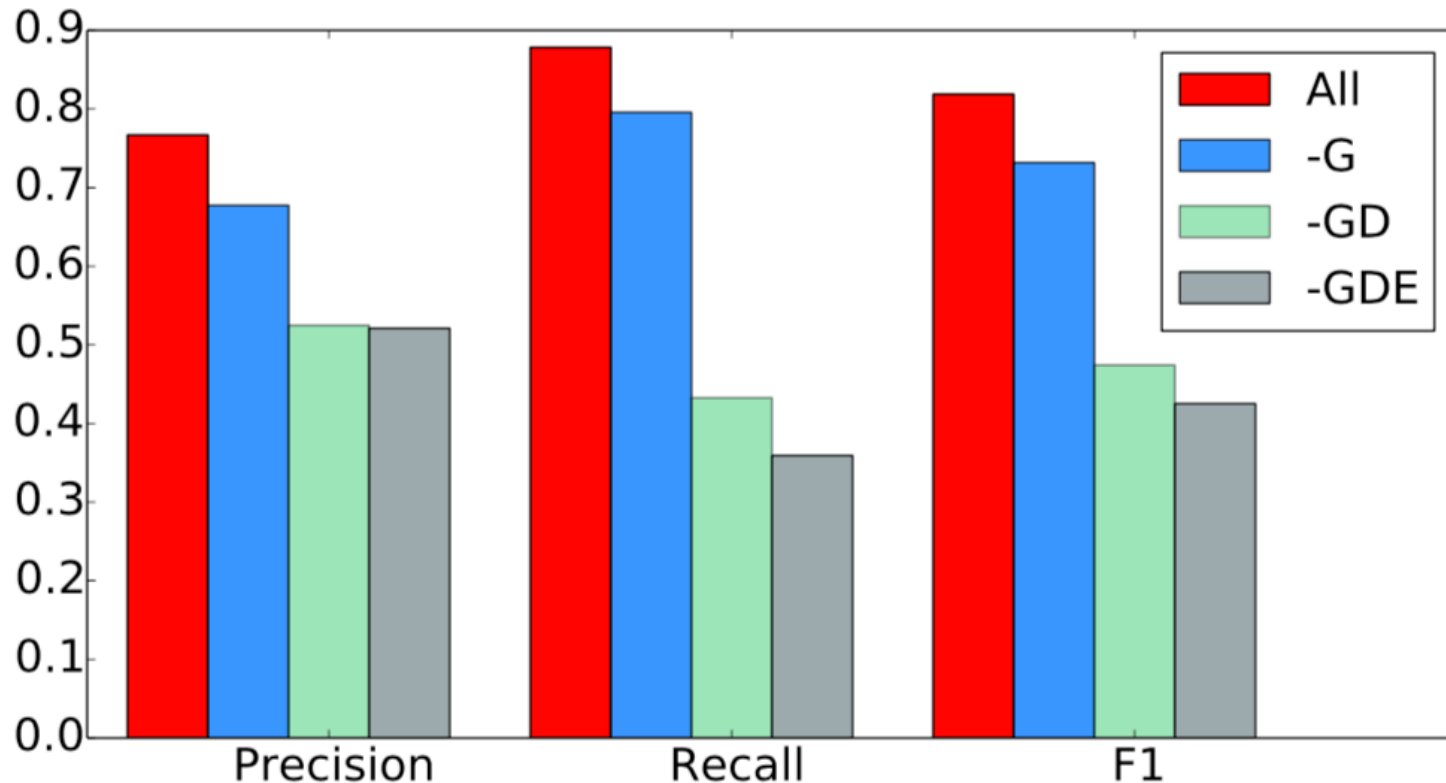
(j) Fraction of townsmen.*

* It is tricky for townsmen, since locals do not have townsmen that are not from Shanghai and always have 0 in this feature.

Experiment Setup

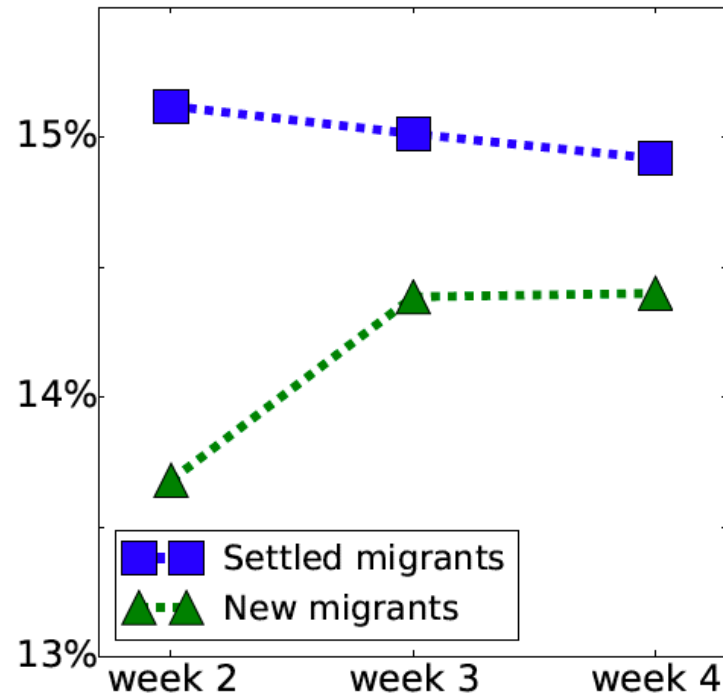
- Training data:
 - 50% of data randomly drawn in week 2
- Testing data:
 - the rest 50% of data in week 2
 - 100% of data in week 3 and week 4
- Classifier:
 - l_2 -regularized logistic regression

Settled Migrants vs. Locals



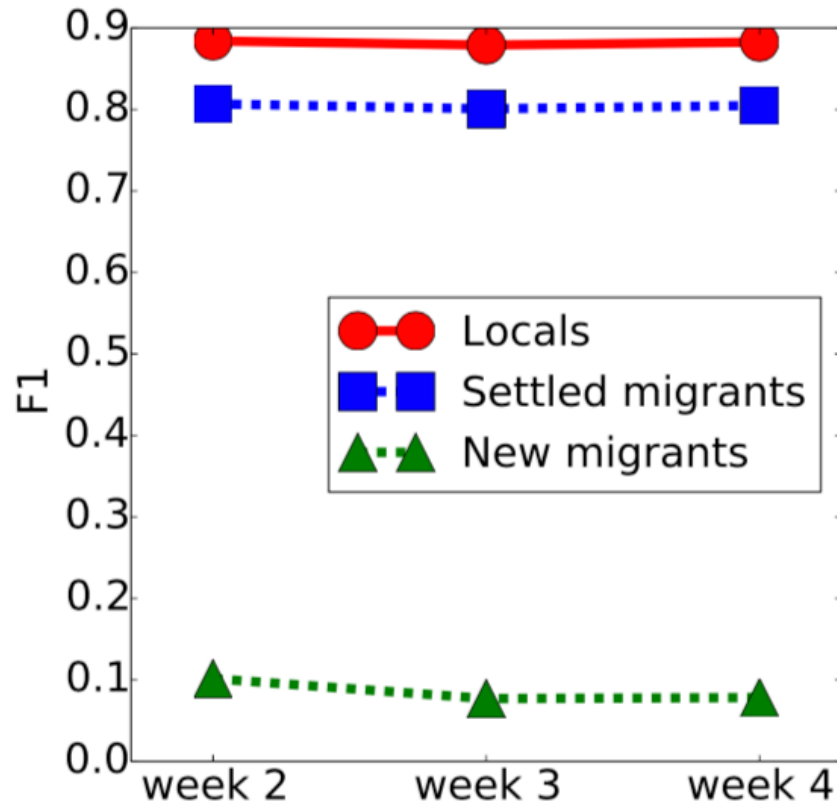
Prediction performance in distinguishing settled migrants in week 2

Integration of New Migrants



Fraction of settled and new migrants
which are misclassified as local

Identify New Migrants



Performance of three-way classification

Conclusion

- We present the first large-scale study on migrant integration based on telecommunication metadata.
- We find that migrants have more diverse contacts and move around the city with a larger radius than locals after they settle down.
- Despite the short time span, we find that migrants are indeed approaching locals in most characteristics.
- We hope that our study can encourage more researchers to examine this important problem!

Conclusion

- We present the first large-scale study on migrant integration based on telecommunication metadata.
- We find that migrants have more diverse contacts and move around the city with a larger radius than locals after they settle down.
- Despite the short time span, we find that migrants are indeed approaching locals in most characteristics.
- We hope that our study can encourage more researchers to examine this important problem!

Thank you!

Q&A