



Smithsonian  
*Cooper-Hewitt, National Design Museum*

AK

# THE TRAILS IN US

PATTERNS OF HUMAN BEHAVIOR



[akamlani@gmail.com](mailto:akamlani@gmail.com) | [arikamlani.com](http://arikamlani.com)



[@akamlani/cooperhewitt](https://github.com/akamlani/cooperhewitt)

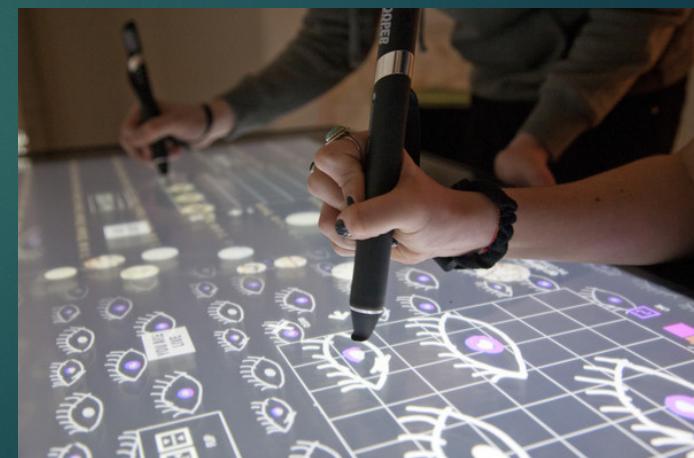
# THE PEN ARRIVES

## WHY THE PEN



AK

- ▶ ENHANCE VISITOR EXPERIENCE
  - ▶ DRAW VISITOR INTO THE DESIGN EXPERIENCE
- ▶ RECENT DATASET RELEASED MAR 2016
  - ▶ 3.5M+ OBSERVATIONS
  - ▶ AGGREGATE W/ COLLECTION METADATA (80K+)
- ▶ FOCUS: VISITOR SEQUENCE PATTERN TAGGING BEHAVIOR



# PAGE RANK



INFLUENTIAL ARTWORKS BASED ON VISITOR'S DAILY JOURNEY



RM 206



RM 212/105



RM 212/105



RM 212/105



RM 206

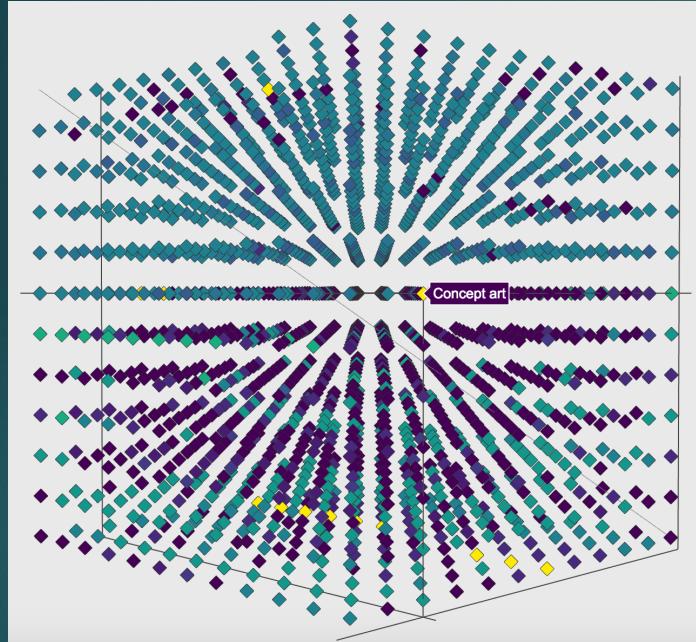


RM 212/105



# COMMUNITIES

## LABEL PROPAGATION ALGORITHM (LPA)



[Interactive Community Detection Demo](#)

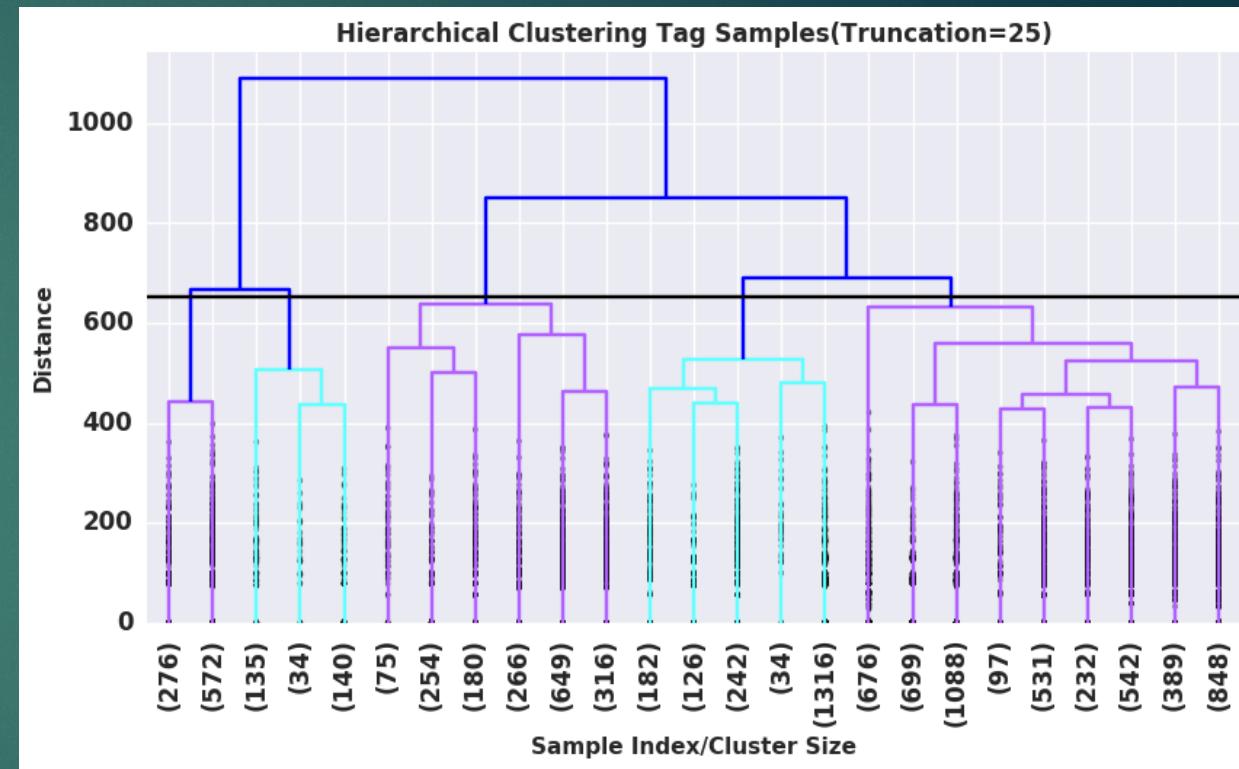


- ▶ ONE LARGE DENSE COMMUNITY
- ▶ ONE SMALL COMMUNITY
- ▶ SEVERAL SMALL MICRO-COMMUNITIES

DENSE	SMALL
EVEN DISTRIBUTION OF EXHIBITIONS	NO EXHIBITIONS
HIGH CAPACITY ROOMS (RM 205,202) THEME ROOMS (1 <sup>st</sup> FLOOR)	SMALL CAPACITY ROOMS SPECIFIC TO 3 <sup>RD</sup> FLOOR
TYPES: DRAWINGS/CONCEPT ART	TYPES: PRINTS
SPECIAL EVENTS	

# TEMPORAL PATTERNS

- ▶ KMEANS NOT GOOD FOR CATEGORICAL DATA
- ▶ HIERARCHICAL CLUSTERING (AGGLOMERATIVE)
  - ▶ METRIC: COPH DISTANCE (CITYBLOCK) ~ 70%
  - ▶ COMPUTATIONALLY EXPENSIVE
    - ▶ NO IMPLEMENTATION IN SPARK MLlib/ML
- ▶ CLUSTERS (PER FEATURE ENGINEERING)
  - ▶ FLOORS (THEMES, LARGE CAPACITY, EXHIBITS)
  - ▶ LOCATIONS (ROOMS/CAPACITY/TURNOVER)
  - ▶ BURSTY VISITORS: (2<sup>ND</sup> FLOOR, LARGE CAPACITY)
  - ▶ SPECIAL EVENTS (AFTER HOURS)



# RECOMMENDATIONS

## HOW TO HELP THE MUSEUM

- ▶ FURTHER IMPROVE ON VISITOR EXPERIENCE – NO MISSED OPPORTUNITIES
- ▶ OPPORTUNITY TO IMPROVE ON TRANSITION SPOT LOCATIONS
- ▶ TAILORED VISITOR SEGMENTATIONS

# FUTURE VISION

- ▶ ADDITIONAL MODELS ON MEMORY SEQUENCES (RNN/LSTM)
- ▶ FEATURE ENGINEERING ON ANNOTATED METADATA (NLP)
- ▶ TRANSFER LEARNING (APPLY TO OTHER DOMAINS)

# DISCUSSION

AK



Website: [arikamlani.com](http://arikamlani.com)

Email: [akamlani@gmail.com](mailto:akamlani@gmail.com)



@akamlani

AK

# Appendix

# COMPONENTS



NON-DISTRIBUTED



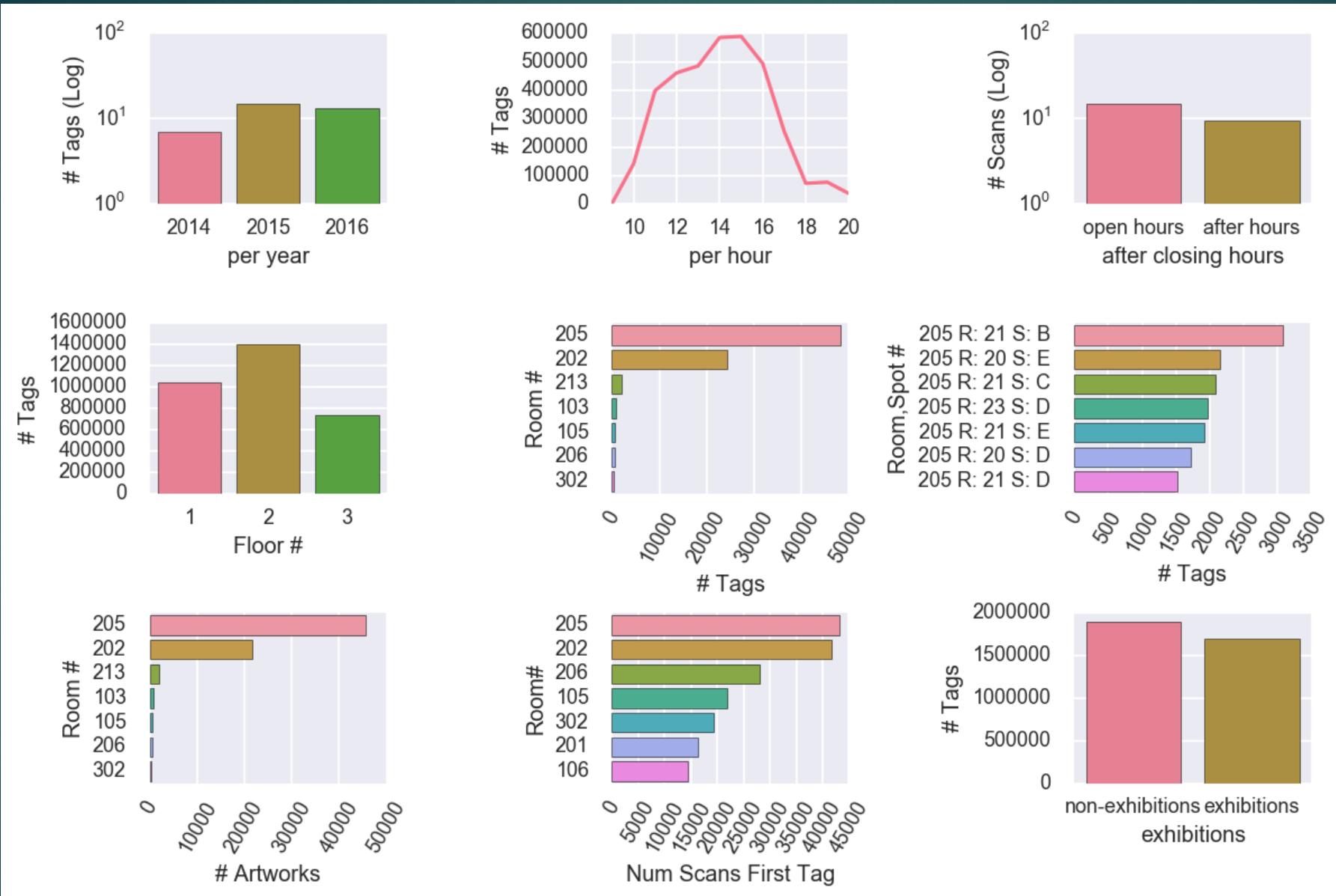
igraph

DISTRIBUTED



# LOVE THY DATA

AK



- ▶ DATA
  - ▶ PEN 3.5M+ OBS
  - ▶ METADATA (80K+)
- ▶ HIGH CAPACITY
  - ▶ RM 205,202
- ▶ FIRST ARRIVAL
  - ▶ RM 205, 206, 202
- ▶ EXHIBITIONS
  - ▶ INFERRED FROM DAILY BUNDLES

# EXHIBITIONS

## ► EXHIBITIONS

- ▶ SMALL CAPACITY ROOMS
  - ▶ FLOOR: 3<sup>RD</sup> NOT 2<sup>ND</sup>
  - ▶ CONSTANT SPOTS
  - ▶ BURSTINESS N/A
  - ▶ METADATA AVAILABLE

