8-11 September 2019, Bled, Slovenia

ADBIS

Contextual and Behavioral Customer Journey Discovery Using a Genetic Approach

Gaël BERNARD

University of Lausanne, Faculty of Business and Economics (HEC), Switzerland

Periklis ANDRITSOS

University of Toronto, Faculty of Information, Canada,

UNIL | Université de Lausanne

UNIVERSITY OF TORONTO

Customer’s interactions are complex and unique

MSI MARKETING SCIENCE INSTITUTE

> *« How does the increasingly complex technological, promotional, and advertising ecosystem affect the customer journey? »* [1]

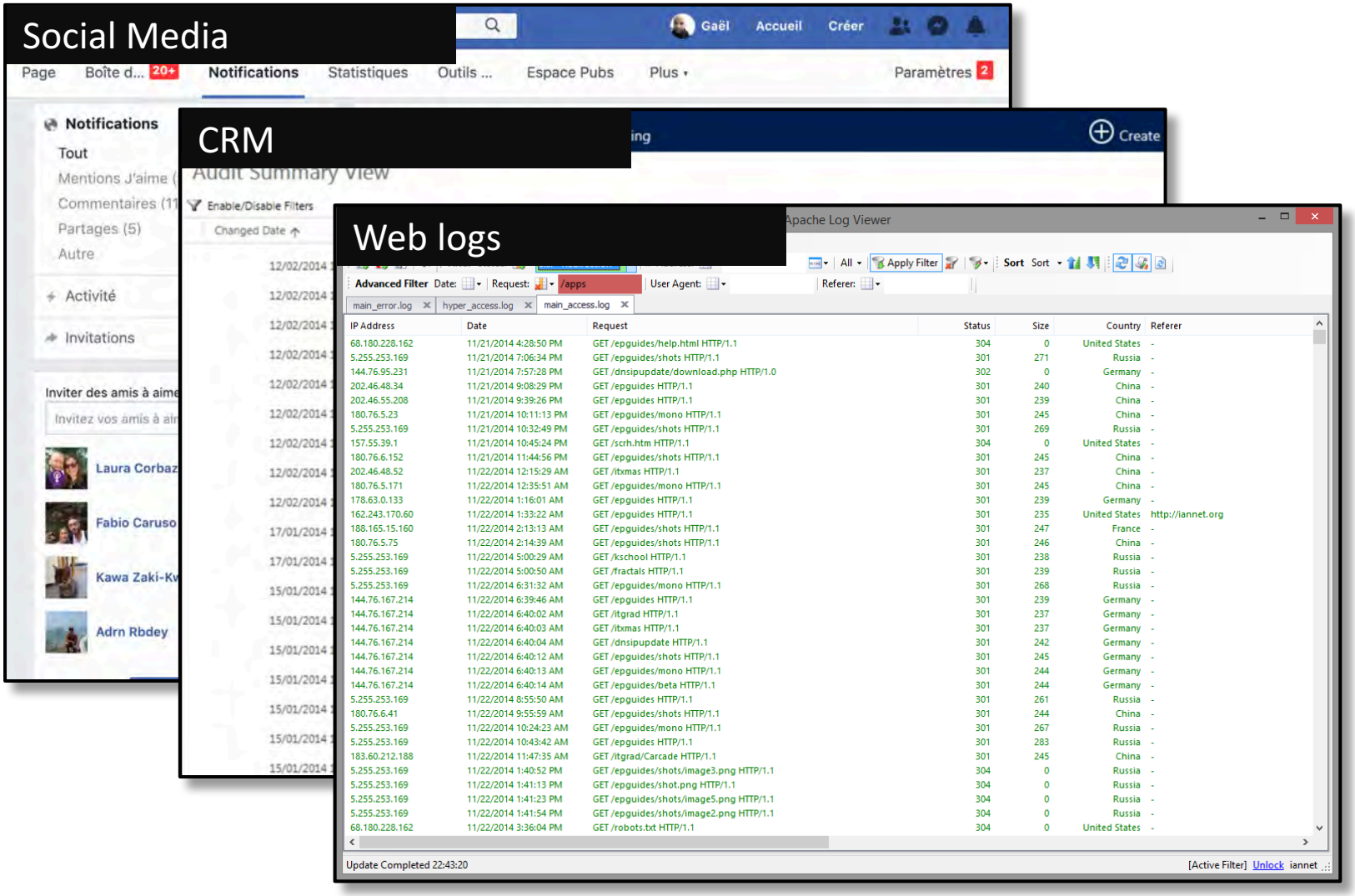
[1] Research priorities 2018-2020. Tech. rep., Marketing Science Institute (2018), https://www.msi.org/research/2018-2020-research-priorities/cultivating-the- customer-asset/1.1.-characterizing-the-customer-journey-along-the-purchase- funnel-and-strategies-to-influence-the-journey/

Customer Journey

- A customer journey contains:
  - List of Touchpoints
    - [looking at product A, buying product B]
  - Contextual data
    - {age:36, gender:male}

- An event log is a collection of customer journeys

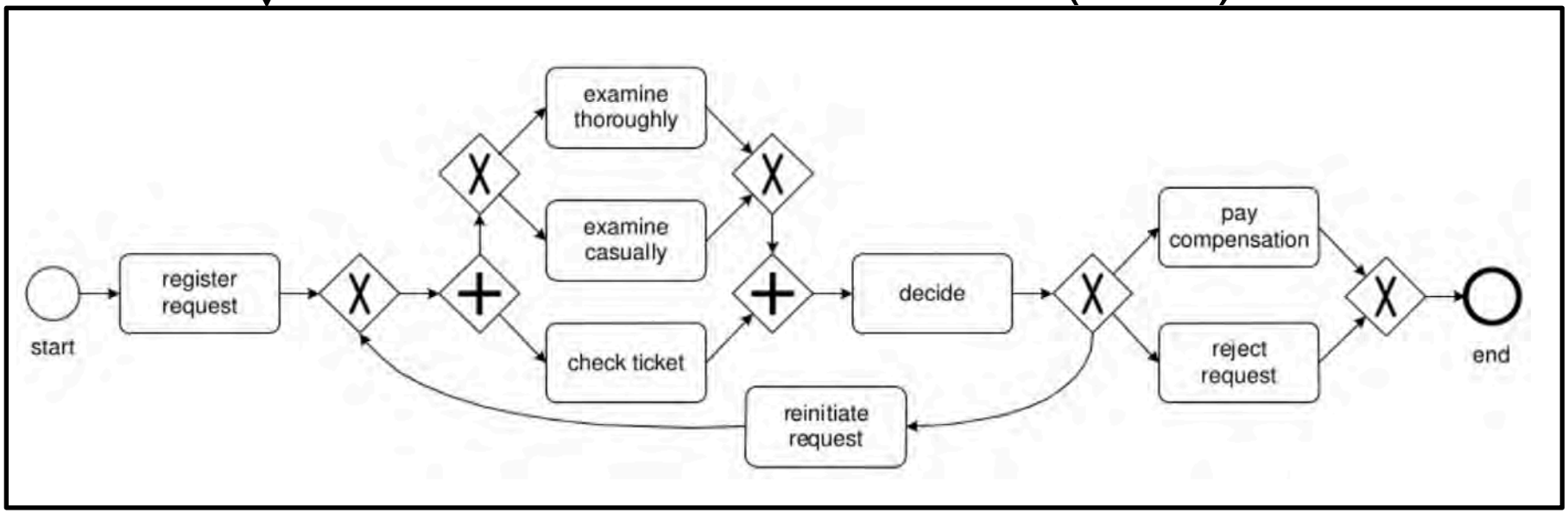
Event logs


Social Media
Gaël
Accueil
Créer
Page
Boîte d... 20+
Notifications
Statistiques
Outils ...
Espace Pubs
Plus
Paramètres 2
Notifications
Tout
Mentions J'aime
Commentaires (11
Partages (5)
Autre
Activité
Invitations
Inviter des amis à aime
Invitez vos amis à air
Laura Corbaz
Fabio Caruso
Kawa Zaki-Kv
Adrn Rbdey
CRM
Audit Summary View
Enable/Disable Filters
Changed Date
Create
Web logs
Apache Log Viewer
Advanced Filter
Date:
Request:
/apps
User Agent:
Referer:
All
Apply Filter
Sort
main_error.log
hyper_access.log
main_access.log
IP Address
Date
Request
Status
Size
Country
Referer
68.180.228.162 11/21/2014 4:28:50 PM GET /epguides/help.html HTTP/1.1 304 0 United States -
5.255.253.169 11/21/2014 7:06:34 PM GET /epguides/shots HTTP/1.1 301 271 Russia -
144.76.95.231 11/21/2014 7:57:28 PM GET /dnsipupdate/download.php HTTP/1.0 302 0 Germany -
202.46.48.34 11/21/2014 9:08:29 PM GET /epguides HTTP/1.1 301 240 China -
202.46.55.208 11/21/2014 9:39:26 PM GET /epguides HTTP/1.1 301 239 China -
180.76.5.23 11/21/2014 10:11:13 PM GET /epguides/mono HTTP/1.1 301 245 China -
5.255.253.169 11/21/2014 10:32:49 PM GET /epguides/shots HTTP/1.1 301 269 Russia -
157.55.39.1 11/21/2014 10:45:24 PM GET /scrh.htm HTTP/1.1 304 0 United States -
180.76.6.152 11/21/2014 11:44:56 PM GET /epguides/shots HTTP/1.1 301 245 China -
202.46.48.52 11/22/2014 12:15:29 AM GET /itxmas HTTP/1.1 301 237 China -
180.76.5.171 11/22/2014 12:35:51 AM GET /epguides/mono HTTP/1.1 301 245 China -
178.63.0.133 11/22/2014 1:16:01 AM GET /epguides HTTP/1.1 301 239 Germany -
162.243.170.60 11/22/2014 1:33:22 AM GET /epguides HTTP/1.1 301 235 United States http://iannet.org
188.165.15.160 11/22/2014 2:13:13 AM GET /epguides/shots HTTP/1.1 301 247 France -
180.76.5.75 11/22/2014 2:14:39 AM GET /epguides/shots HTTP/1.1 301 246 China -
5.255.253.169 11/22/2014 5:00:29 AM GET /kschool HTTP/1.1 301 238 Russia -
5.255.253.169 11/22/2014 5:00:50 AM GET /fractals HTTP/1.1 301 239 Russia -
5.255.253.169 11/22/2014 6:31:32 AM GET /epguides/mono HTTP/1.1 301 268 Russia -
144.76.167.214 11/22/2014 6:39:46 AM GET /epguides HTTP/1.1 301 239 Germany -
144.76.167.214 11/22/2014 6:40:02 AM GET /itgrad HTTP/1.1 301 237 Germany -
144.76.167.214 11/22/2014 6:40:03 AM GET /itxmas HTTP/1.1 301 237 Germany -
144.76.167.214 11/22/2014 6:40:04 AM GET /dnsipupdate HTTP/1.1 301 242 Germany -
144.76.167.214 11/22/2014 6:40:12 AM GET /epguides/shots HTTP/1.1 301 245 Germany -
144.76.167.214 11/22/2014 6:40:13 AM GET /epguides/mono HTTP/1.1 301 244 Germany -
144.76.167.214 11/22/2014 6:40:14 AM GET /epguides/beta HTTP/1.1 301 244 Germany -
5.255.253.169 11/22/2014 8:55:50 AM GET /epguides HTTP/1.1 301 261 Russia -
180.76.6.41 11/22/2014 9:55:59 AM GET /epguides/shots HTTP/1.1 301 244 China -
5.255.253.169 11/22/2014 10:24:23 AM GET /epguides/mono HTTP/1.1 301 267 Russia -
5.255.253.169 11/22/2014 10:43:42 AM GET /epguides HTTP/1.1 301 283 Russia -
183.60.212.188 11/22/2014 11:47:35 AM GET /itgrad/Carcade HTTP/1.1 301 245 China -
5.255.253.169 11/22/2014 1:40:52 PM GET /epguides/shots/image3.png HTTP/1.1 304 0 Russia -
5.255.253.169 11/22/2014 1:41:13 PM GET /epguides/shot.png HTTP/1.1 304 0 Russia -
5.255.253.169 11/22/2014 1:41:23 PM GET /epguides/shots/image5.png HTTP/1.1 304 0 Russia -
5.255.253.169 11/22/2014 1:41:54 PM GET /epguides/shots/image2.png HTTP/1.1 304 0 Russia -
68.180.228.162 11/22/2014 3:36:04 PM GET /robots.txt HTTP/1.1 304 0 United States -
Update Completed 22:43:20
[Active Filter] Unlock iannet

Process Discovery Algorithm [1]

Event Logs

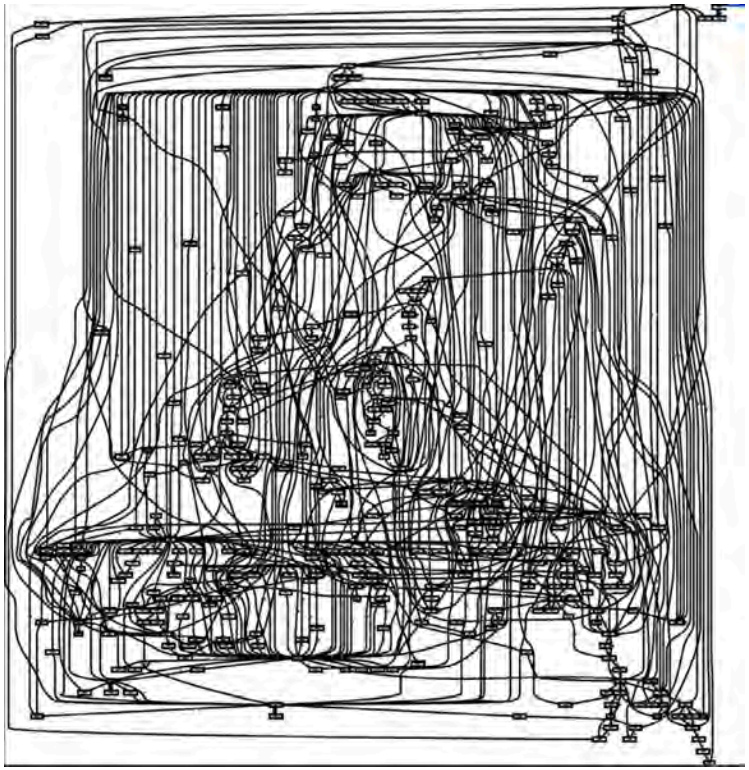
Business Process Model (BPM)



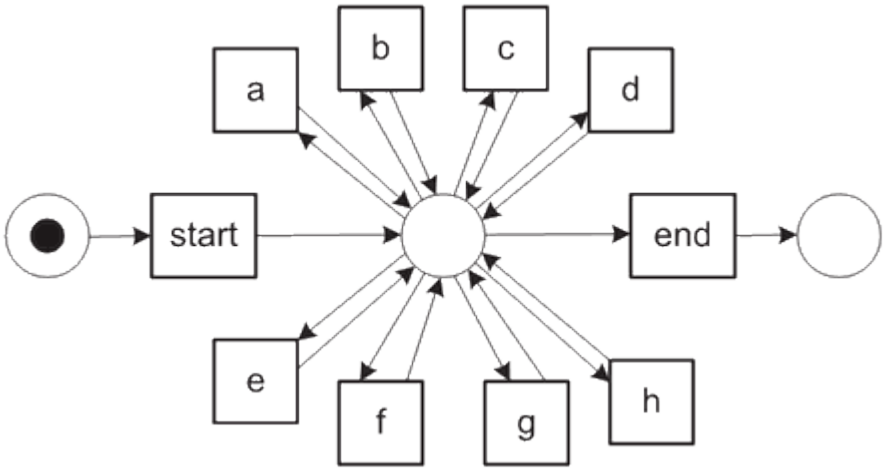
[1] van der Aalst, W.: Process Mining: Data Science in Action. Springer (2016)

BPM on Complex Event Logs

« Spaghetti »

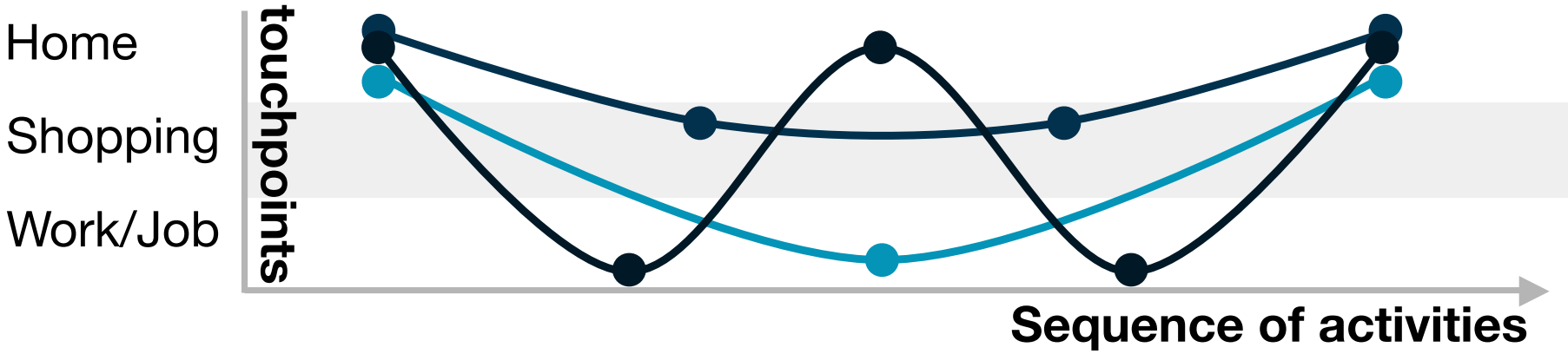


« Flower model »



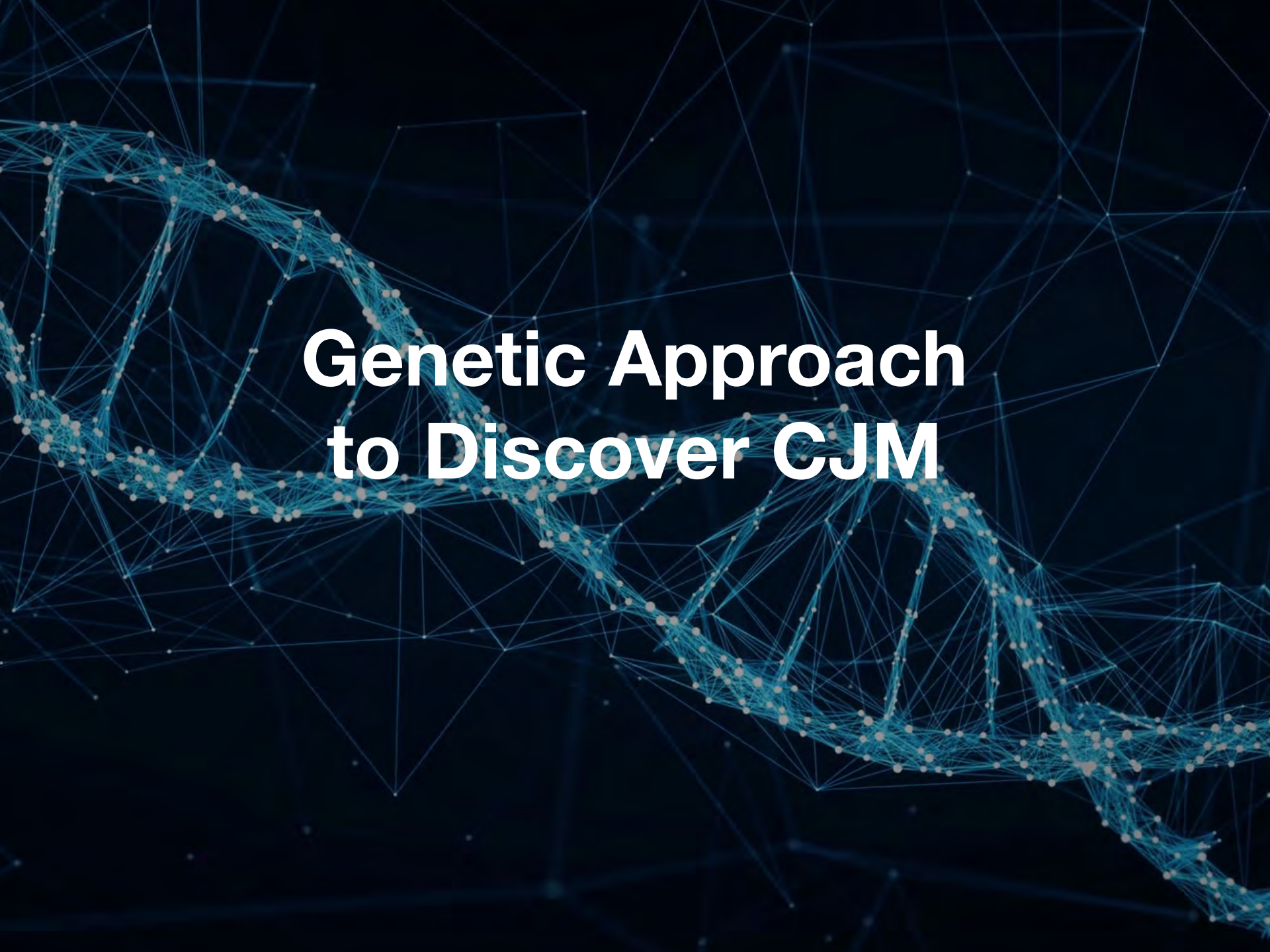
Customer Journey Map (CJM)

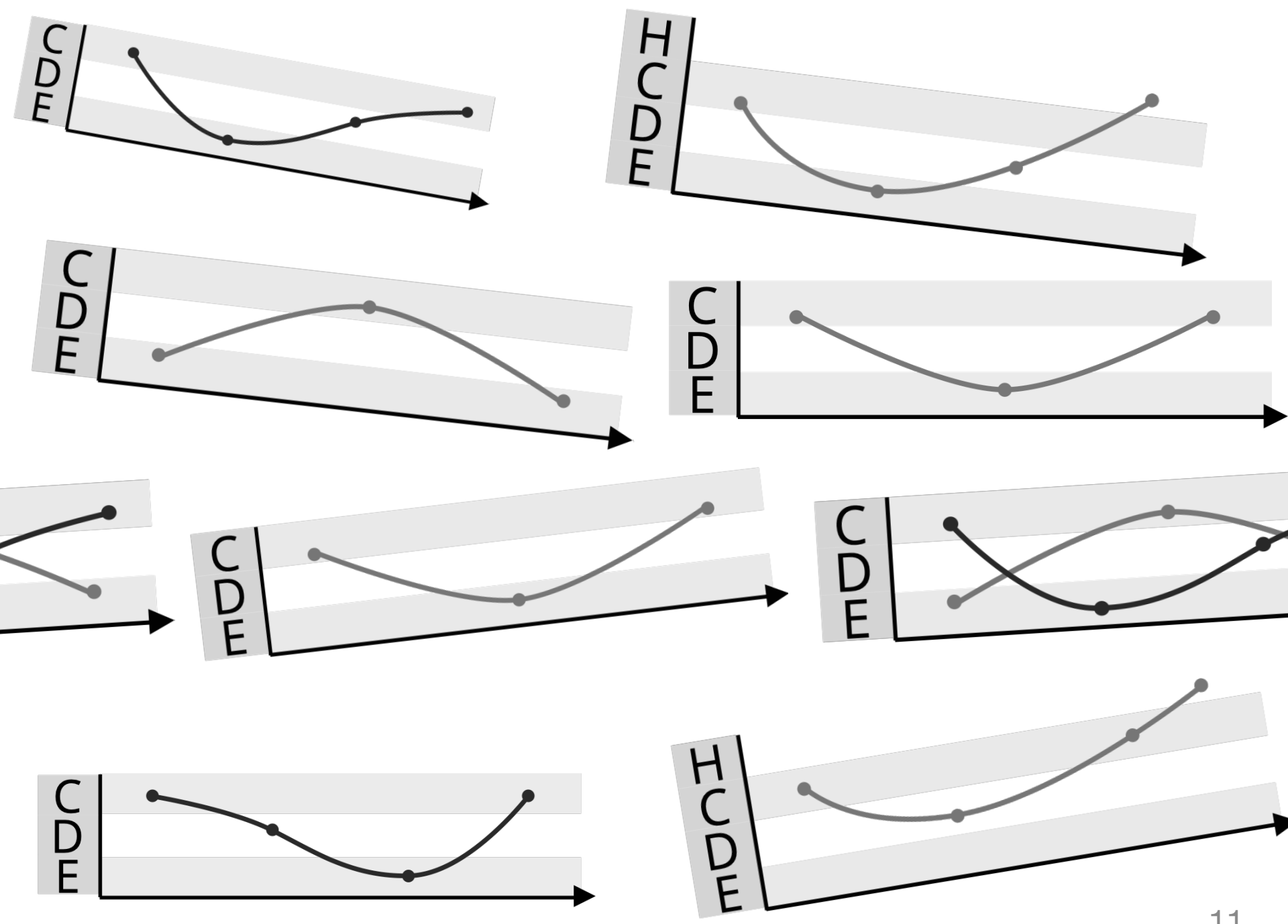
A CJM shows typical journeys experienced by customers

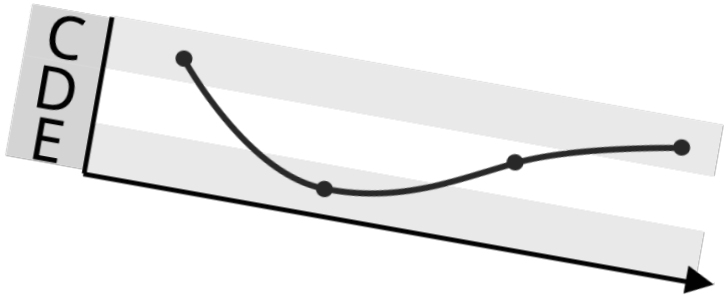


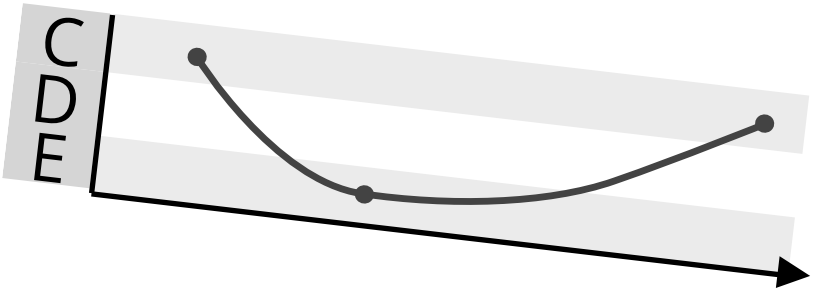
Customer Journey Discovery:

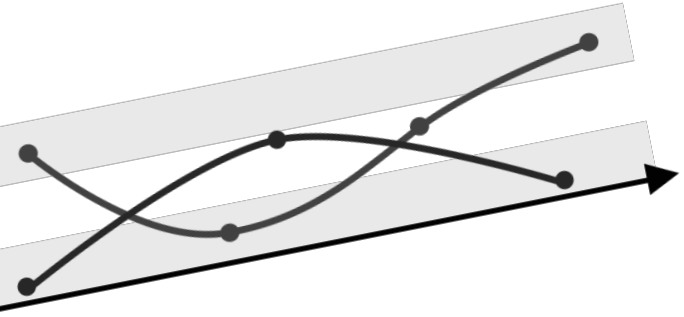
> Finding a *reasonable amount*
> of *representative journeys*
> that *summarize well*
> the actual journeys.

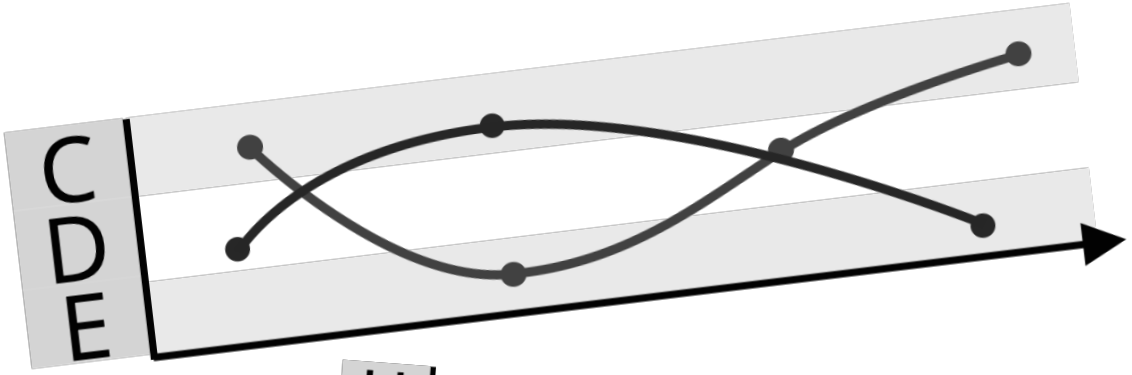

Genetic Approach
to Discover CJM

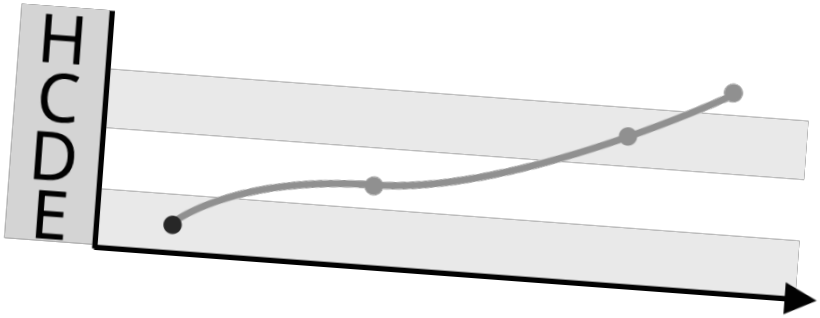

C
D
E
H
C
D
E
C
D
E
C
D
E
C
D
E
C
D
E
C
D
E
H
C
D
E

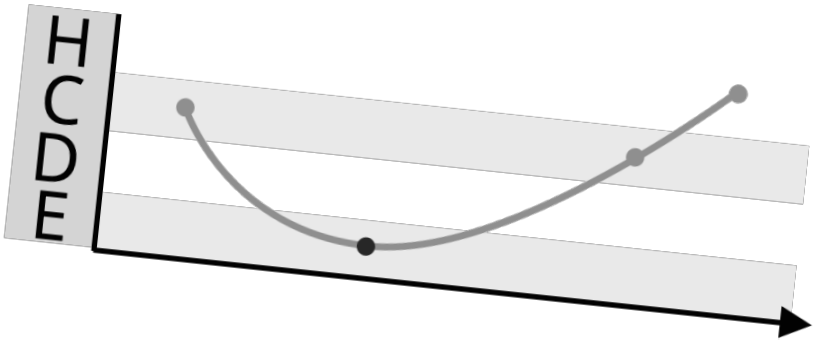

C
D
E

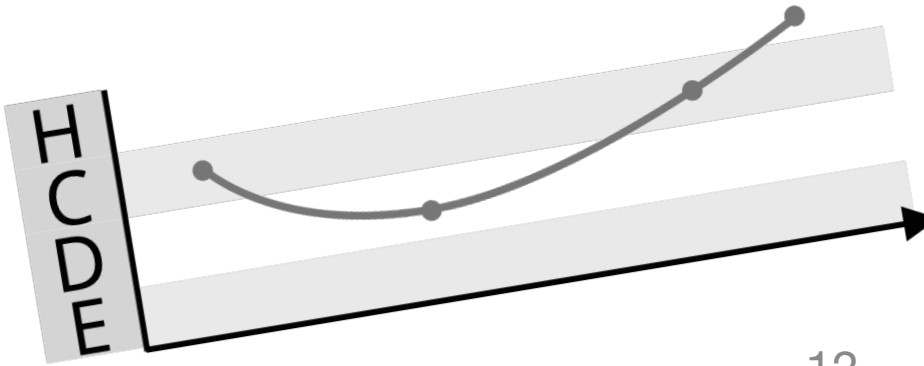

C
D
E

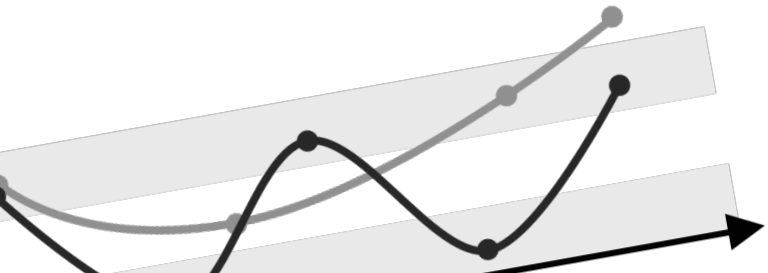


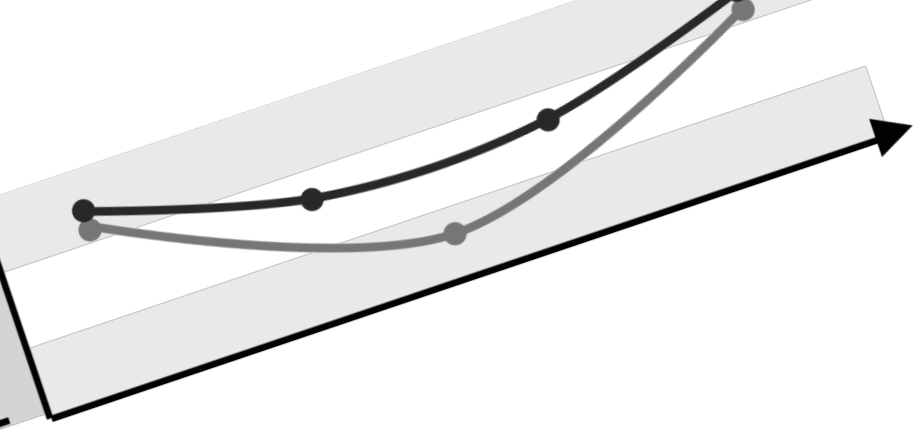

C
D
E

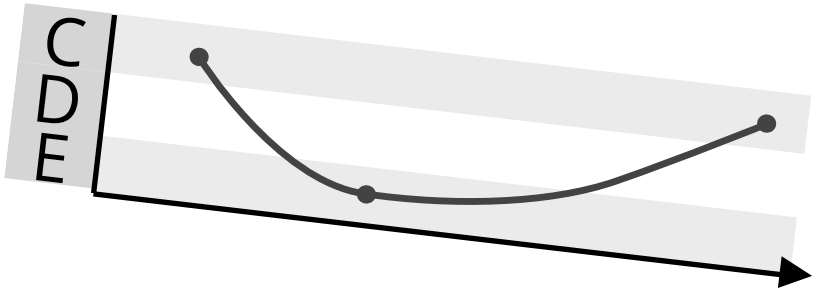

H
C
D
E

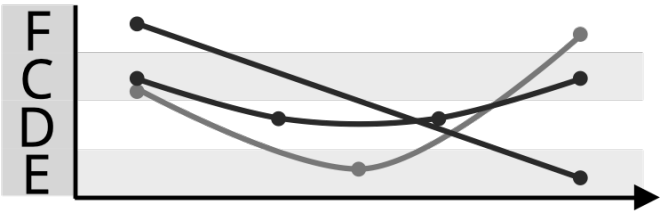

H
C
D
E

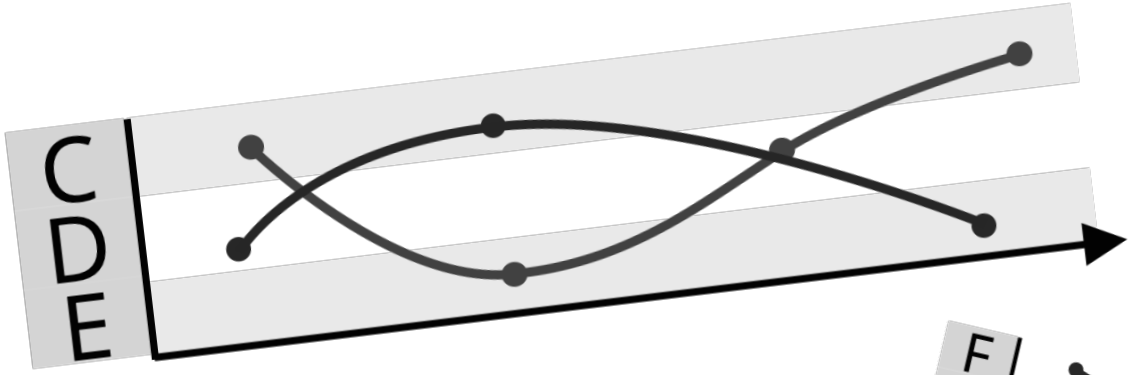

H
C
D
E

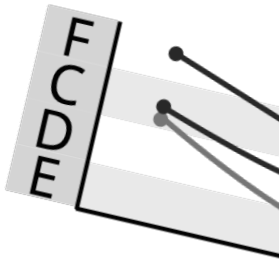


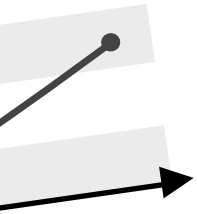


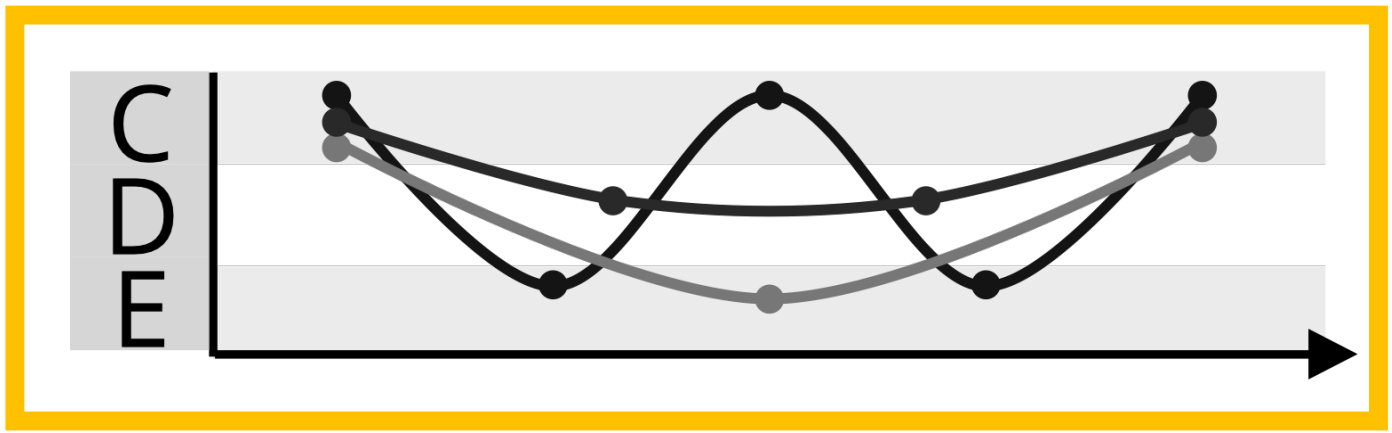

C
D
E


F
C
D
E


C
D
E

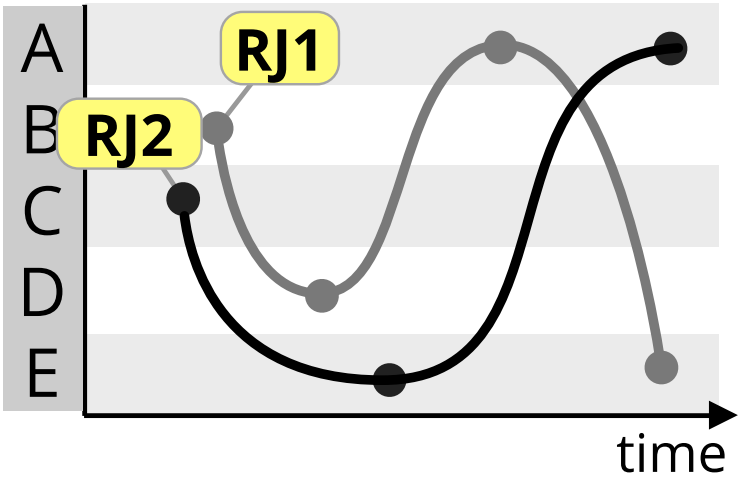

F
C
D
E




C
D
E

Assign

Assigned actual journeys to the closest representative using the Levenshtein distance



Actual Journeys:

RJ1 1, 3, 4, 5,…

RJ2 2, 6, 7, 8,…

Metrics

- Three metrics:
  1. Fitness
  2. Distance from Target
  3. Contextual Distance

$$Fitness(J_a, J_{\mathcal{R}}) = 1 - \frac{\sum_{i=1}^{|J_a|} min_{j=1}^{|J_{\mathcal{R}}|}(Levenshtein(\sigma_{\mathcal{A}_\mathbf{i}}; \sigma_{\mathcal{R}_\mathbf{j}}))}{\sum_{i=1}^{|J_a|} Length(\sigma_{\mathcal{A}_i})}$$

$$DistanceFromTarget\,(k_R, k_h, x_o) = \frac{1}{1 + \left(\frac{|k_{\mathcal{R}} - k_h|}{x_0}\right)^2}$$

Target = k_h, best Calinski-Harabasz Index from 2 to 15 [1]

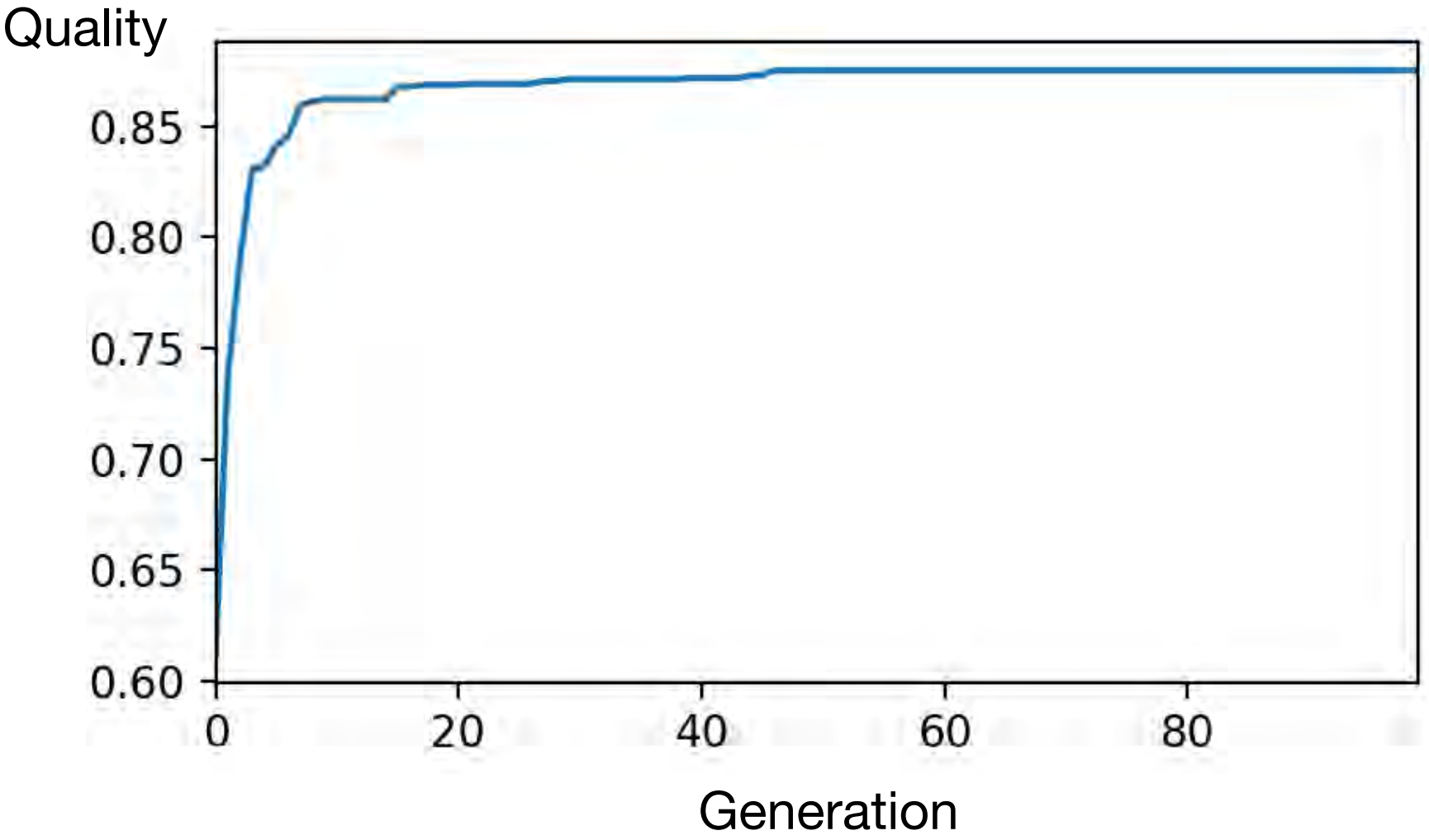
[1] Calinski, T., Harabasz, J.: A dendrite method for cluster analysis. Communications in Statistics-theory and Methods 3(1), 1–27 (1974)

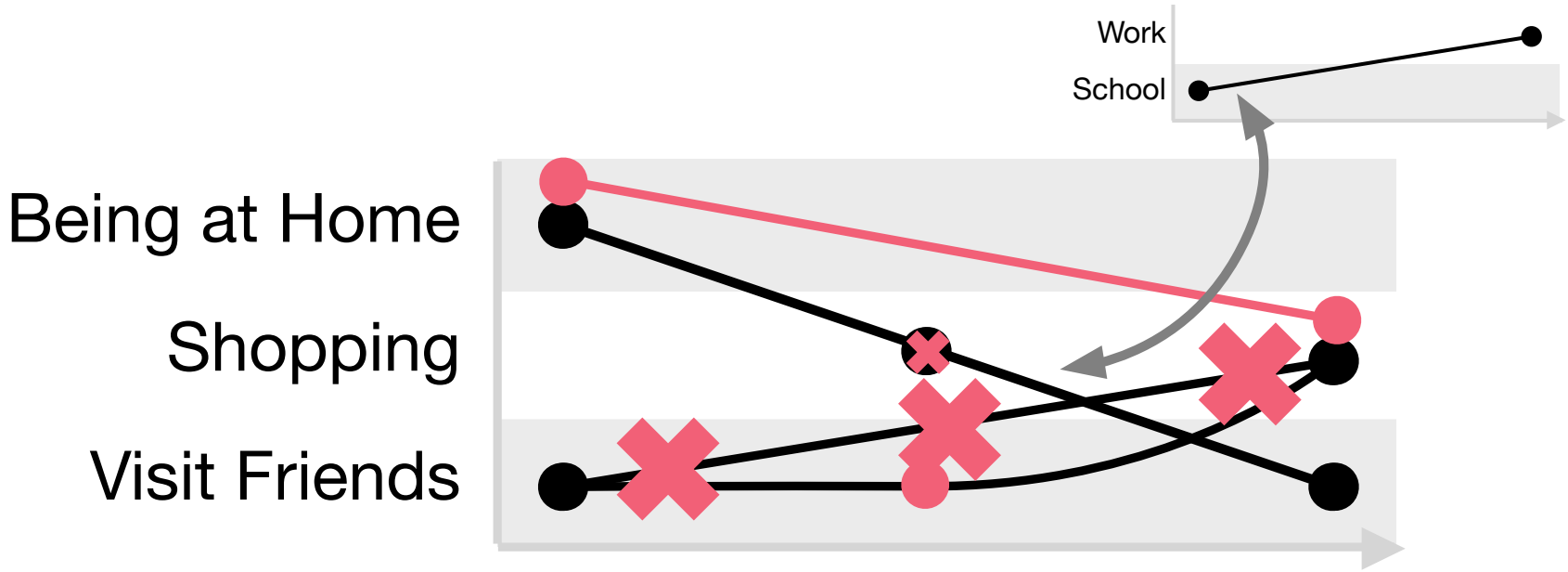
(COSINE)

$$ContextualDistance(v_1, v_2) = \frac{v_1 \cdot v_2}{||v_1|| \cdot ||v_2||}$$

Metrics

- Overall quality is the weighted mean of
  1. Fitness
  2. Distance from Target
  3. Contextual Distance

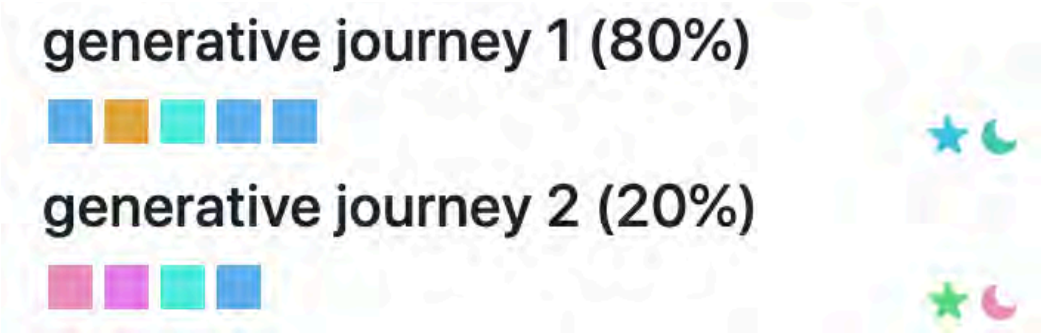

Quality
0.85
0.80
0.75
0.70
0.65
0.60
0
20
40
60
80
Generation



- Add a touchpoint
- Add a journey
- Crossover
- Remove a touchpoint
- Remove a journey

Evaluation: Synthetic Dataset

- Ground truth: Generative Journey



- Synthetic datasets: 40 CJMs of various complexities available online [1]

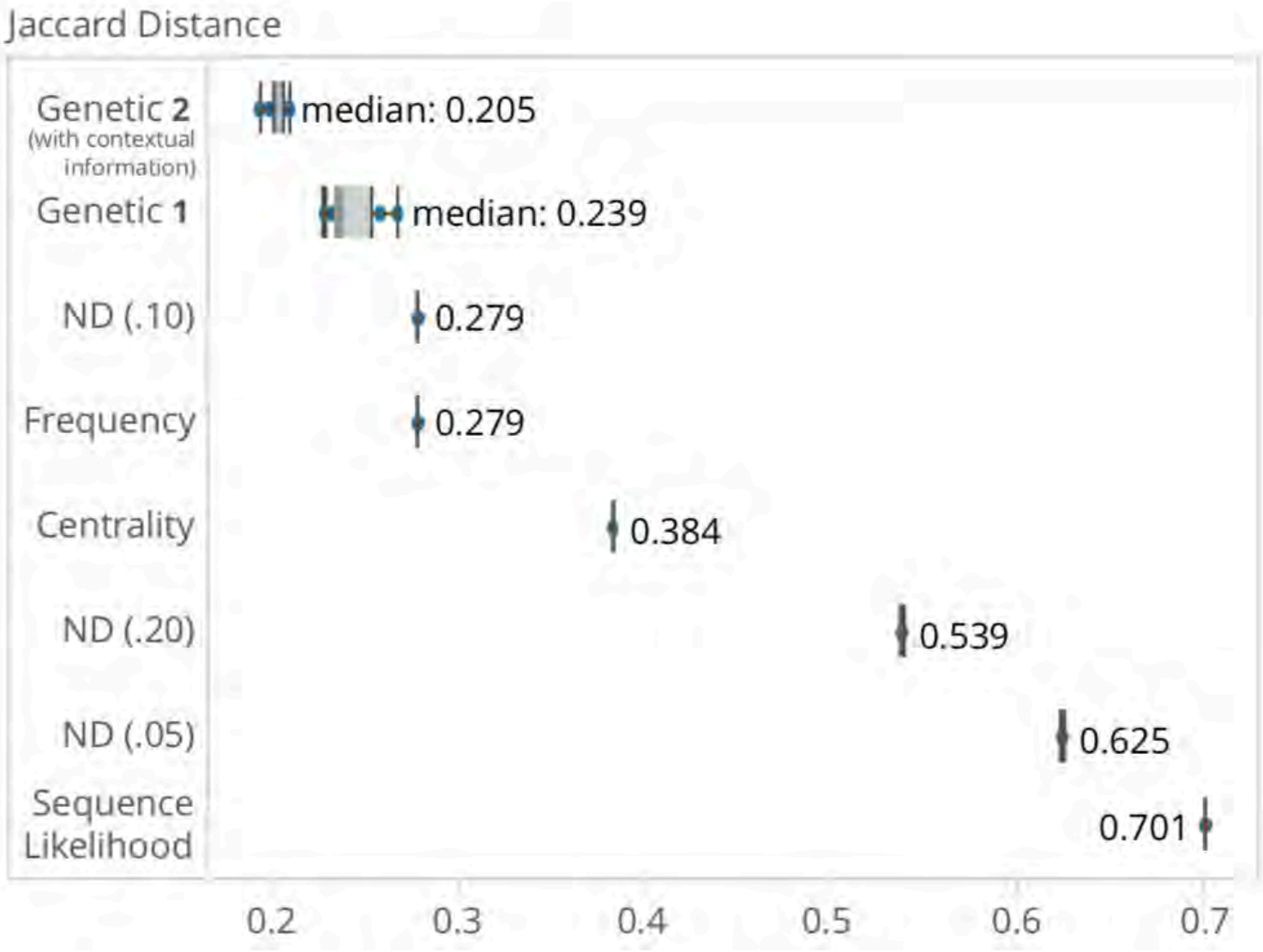
[1] http://customer-journey.unil.ch/datasets/

Evaluation: Setting

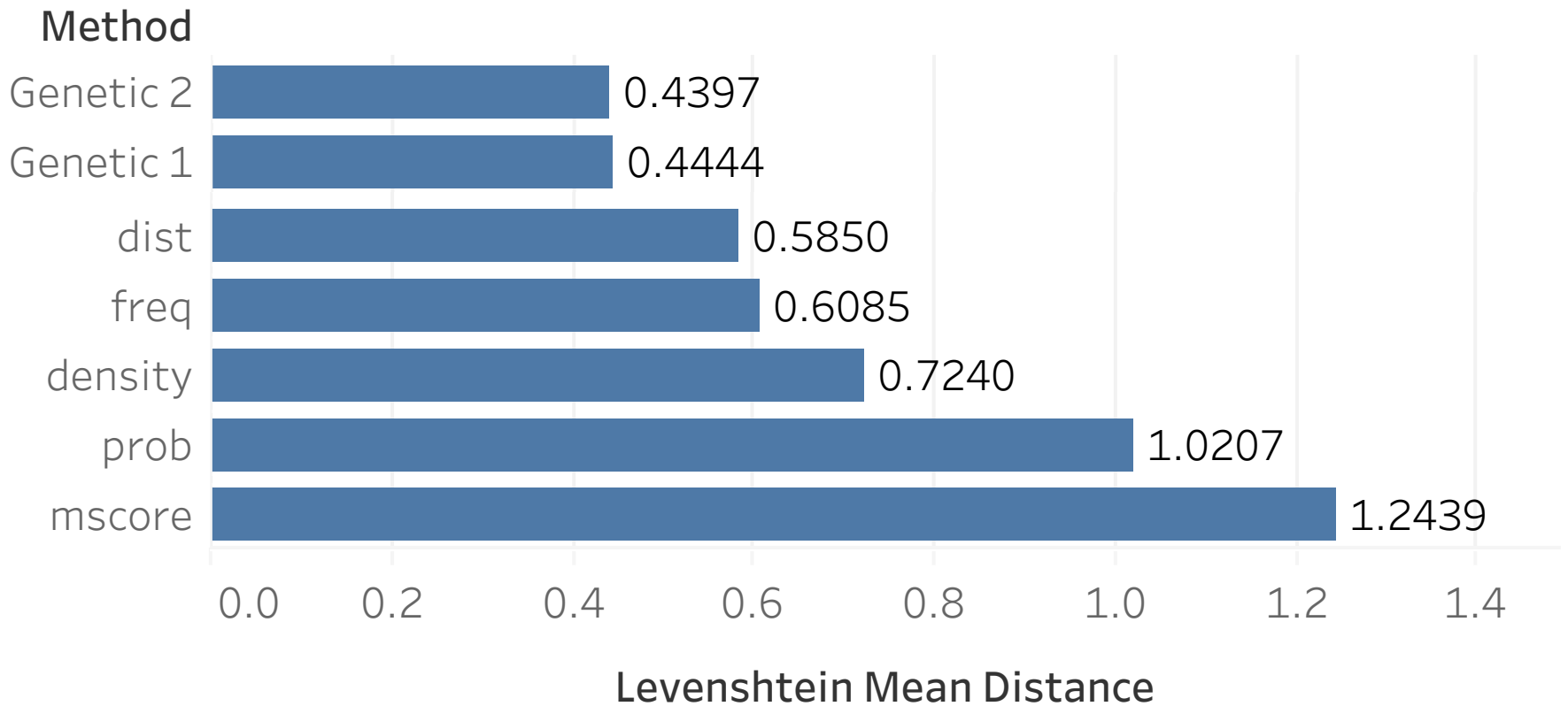
- Compare with Traminer [1]:
  - « Data-driven methods that search for the typical patterns among the observed sequences »

[1] Gabadinho, A. & Ritschard, G. (2013), "Searching for typical life trajectories applied to childbirth histories", In Levy, R. & Widmer, E. (eds)*Gendered life courses - Between individualization and standardization. A European approach applied to Switzerland*, pp. 287-312. Vienna: LIT.Available here.

Evaluation: External Evaluation



Evaluation: Internal Evaluation



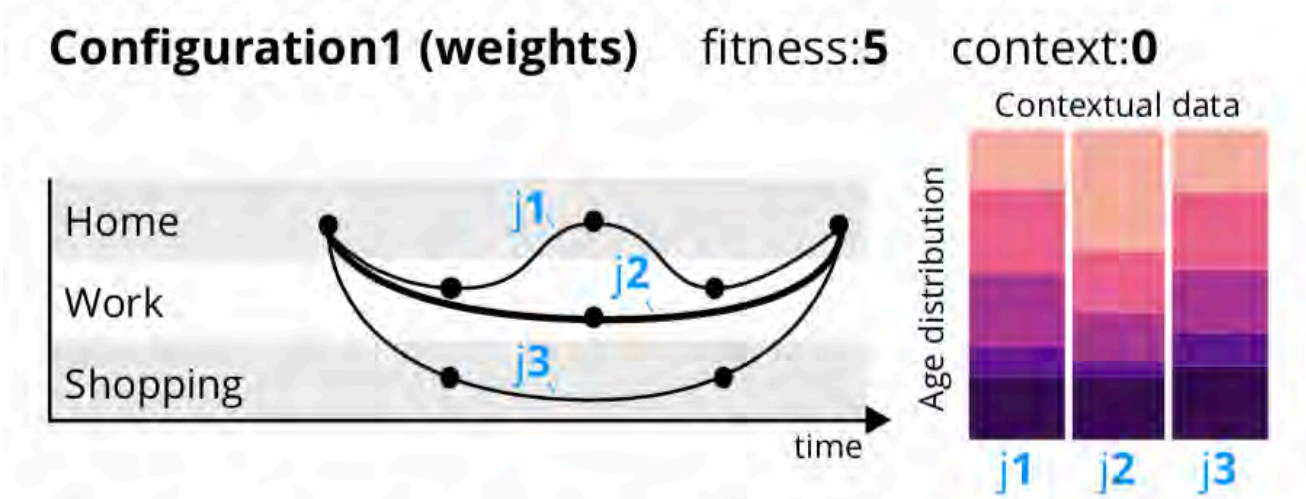
Chicago Dataset

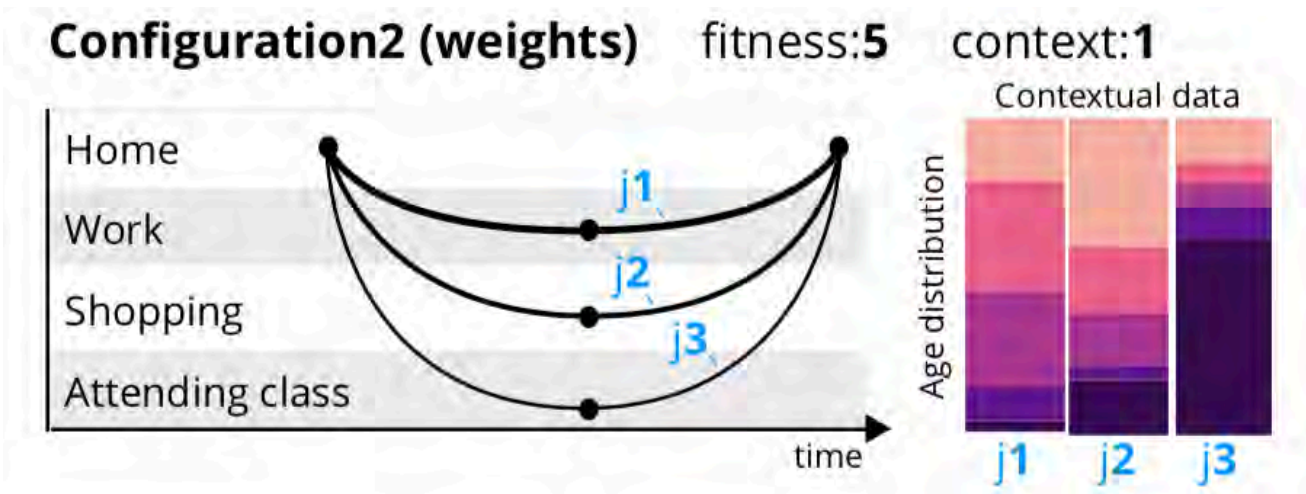
- Publicly available [1]
- Typical journeys from citizens
- Contextual information available: e.g., Age
- ~30K Journeys
- 123K Activities
- 4.1 Avg. activities / journeys

[1]: https://www.cmap.illinois.gov/data/transportation/travel-survey

Results


Context: Citizen's age group
0-16yo
17-29yo
30-45yo
45-59yo
60yo+


Configuration1 (weights)
fitness:5
context:0
Home
Work
Shopping
j1
j2
j3
time
Contextual data
Age distribution
j1
j2
j3


Configuration2 (weights)
fitness:5
context:1
Home
Work
Shopping
Attending class
j1
j2
j3
time
Contextual data
Age distribution
j1
j2
j3

Conclusion

- CJM discovery inspired by process mining
- Domain-agnostic
  - Might be expended for Industry-specific needs
- Limitation:
  - Levenshtein distance is expensive!