Robot Assistance at Home

Seth Teller & Matthew Walter
CS & AI Lab, MIT

May 10, 2013

Robots That Work With People

- Robots are moving into human environments
  - Homes
  - Hospitals
  - Workplaces
- Interactions must be intuitive and safe

Time

Blade cluster
(40 cores)

8 cores

Now: People Accommodating Robots

Courtesy: AeroVironment

Courtesy: US Army

Where We Need to Be

Courtesy: Kinova Robotics

Representational Divide

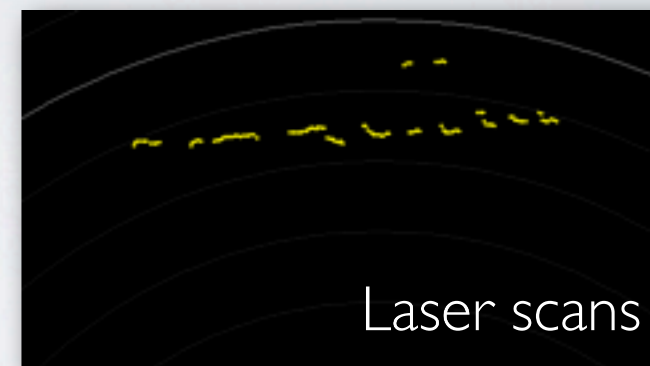
A robot's view of the world is very different from our own

People

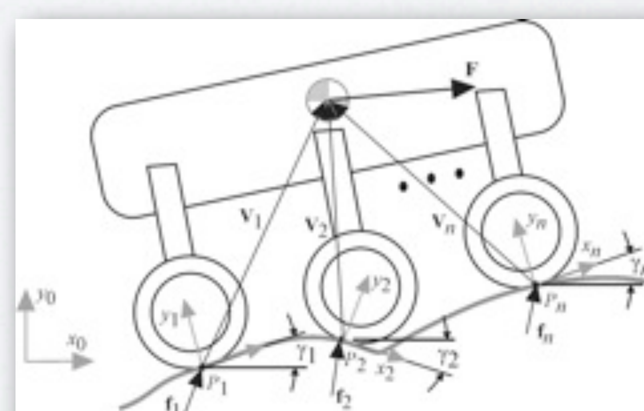
- Objects
- **Places**
- Actions
- People
- Events

Robots

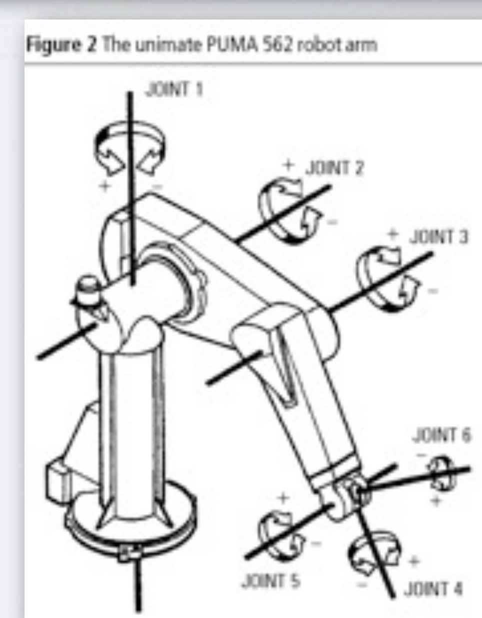
Images



Laser scans



Wheel torques



Joint angles

Learning Semantic Maps from Natural Language

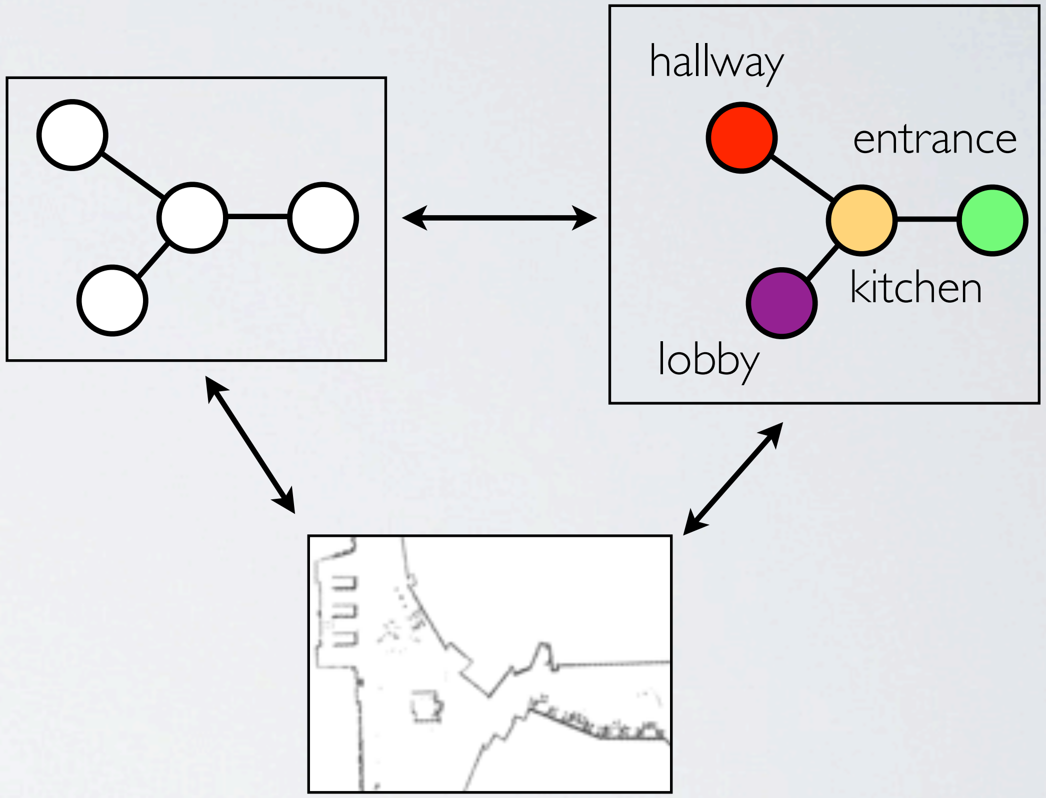
- Human-centric representations of space
  - Spatial relations
  - Semantic attributes (e.g., names, use)
  - Connectivity

- Learn knowledge representation from narrated tour
- Challenges:
  - People convey high-level concepts but robot perception is low-level
  - Spoken descriptions are ambiguous

[RSS'13 (to appear)]

Building Semantic Maps with Natural Language

- Solution:
  - Joint metric, topologic, & semantic model supports information fusion
  - Efficient inference strategy
  - Enable layers to influence one another
  - Utilize natural language grounding framework



[RSS'13 (to appear)]

Model: Posterior over Semantic Graphs

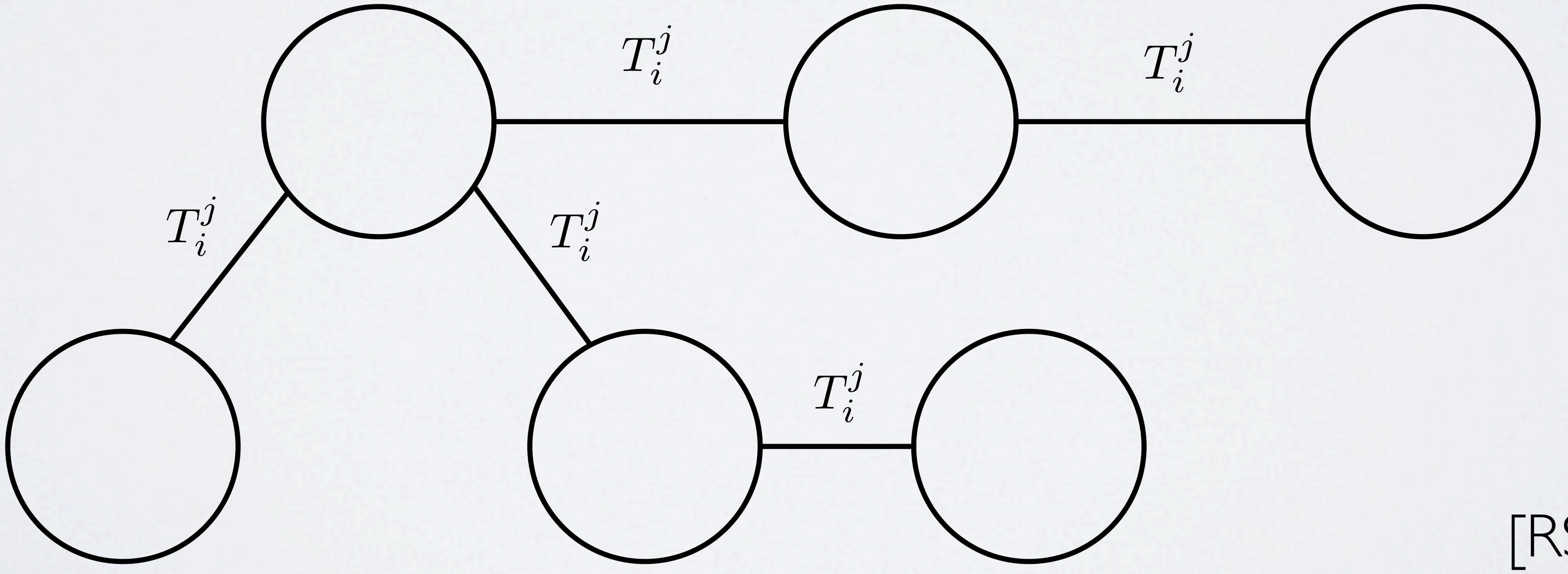
$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t)$$

[RSS'13 (to appear)]

Model: Posterior over Semantic Graphs

$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t)$$

Topology $G_t = (V_t, E_t)$



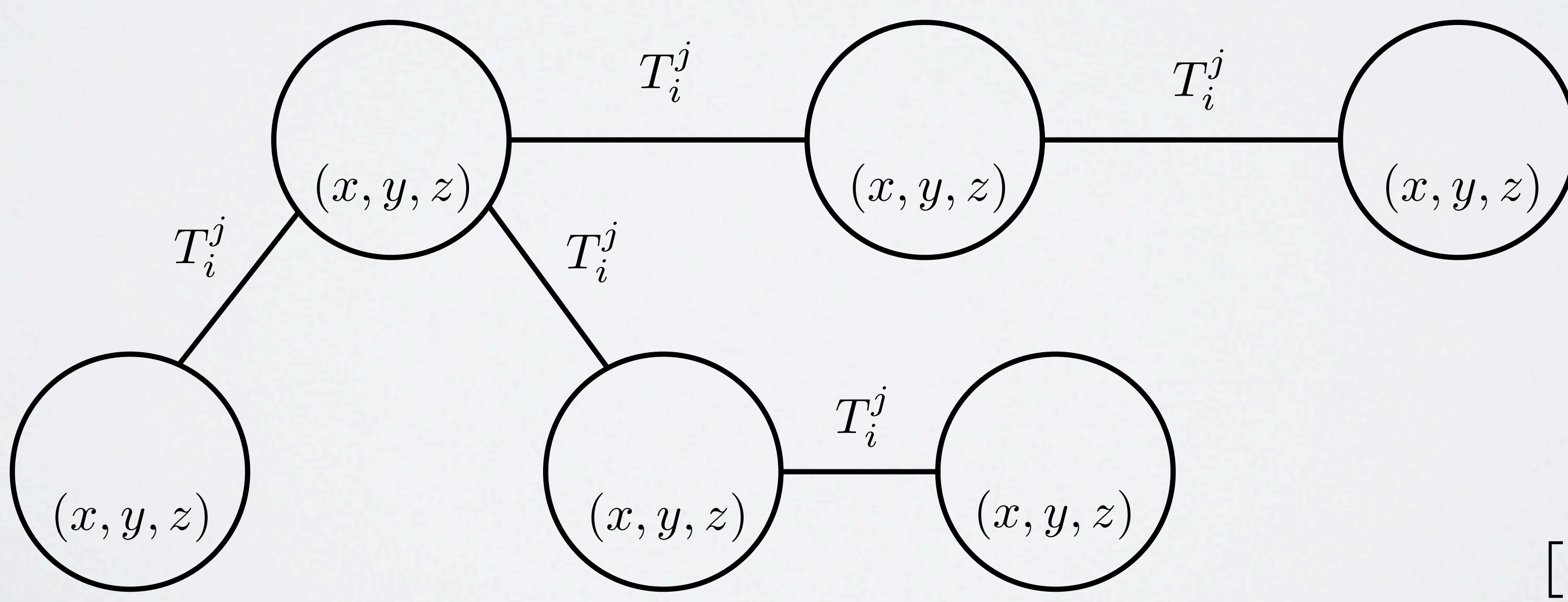
[RSS'13 (to appear)]

Model: Posterior over Semantic Graphs

$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t)$$

Topology $G_t = (V_t, E_t)$

Vertex poses



[RSS'13 (to appear)]

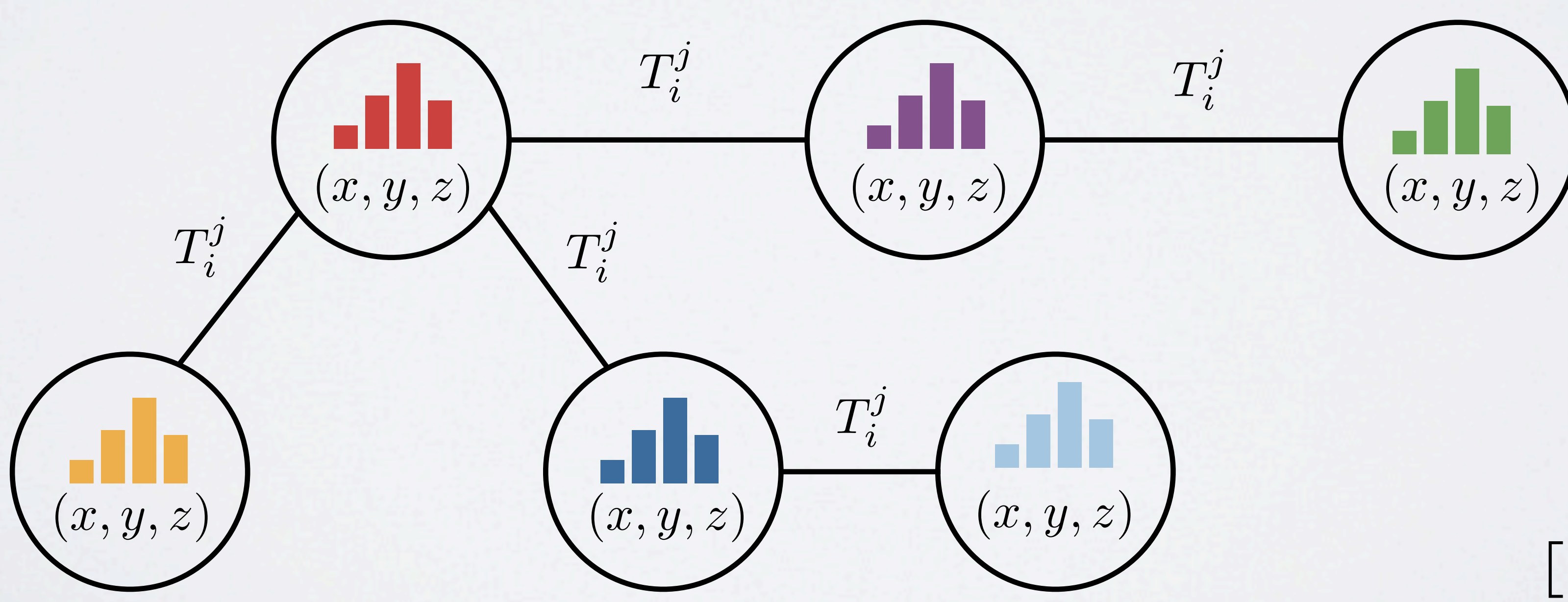
Model: Posterior over Semantic Graphs

$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t)$$

Topology $G_t = (V_t, E_t)$

Semantic labels

Vertex poses



[RSS'13 (to appear)]

Model: Posterior over Semantic Graphs

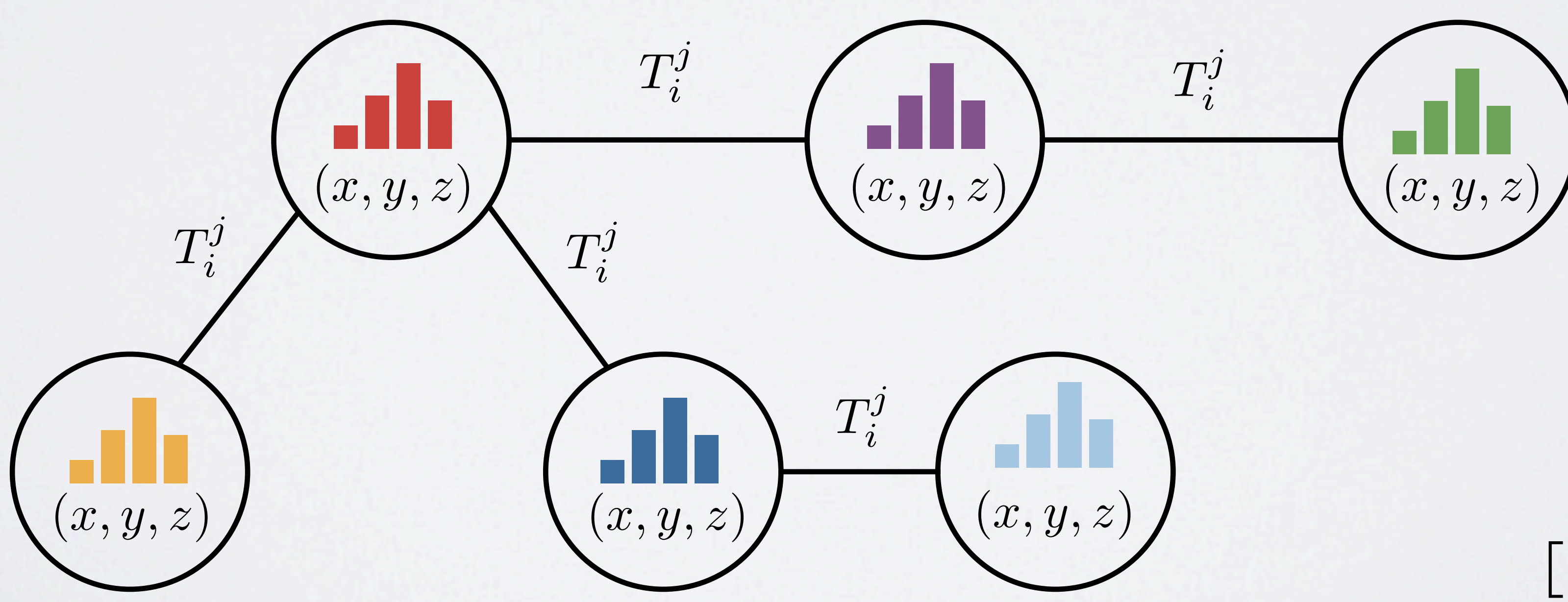
Sensor stream

$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t)$$

Topology $G_t = (V_t, E_t)$

Vertex poses

Semantic labels



[RSS'13 (to appear)]

Model: Posterior over Semantic Graphs

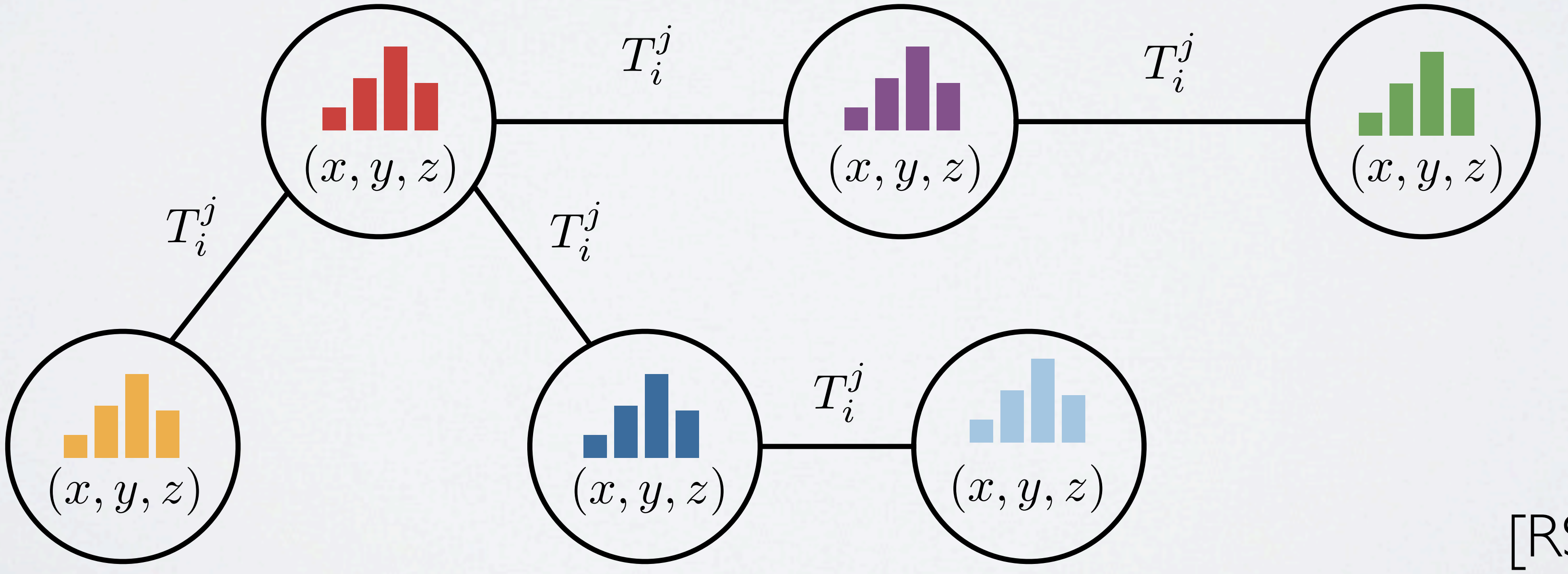
Sensor stream   Odometry

$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t)$$

Topology $G_t = (V_t, E_t)$

Vertex poses

Semantic labels



[RSS'13 (to appear)]

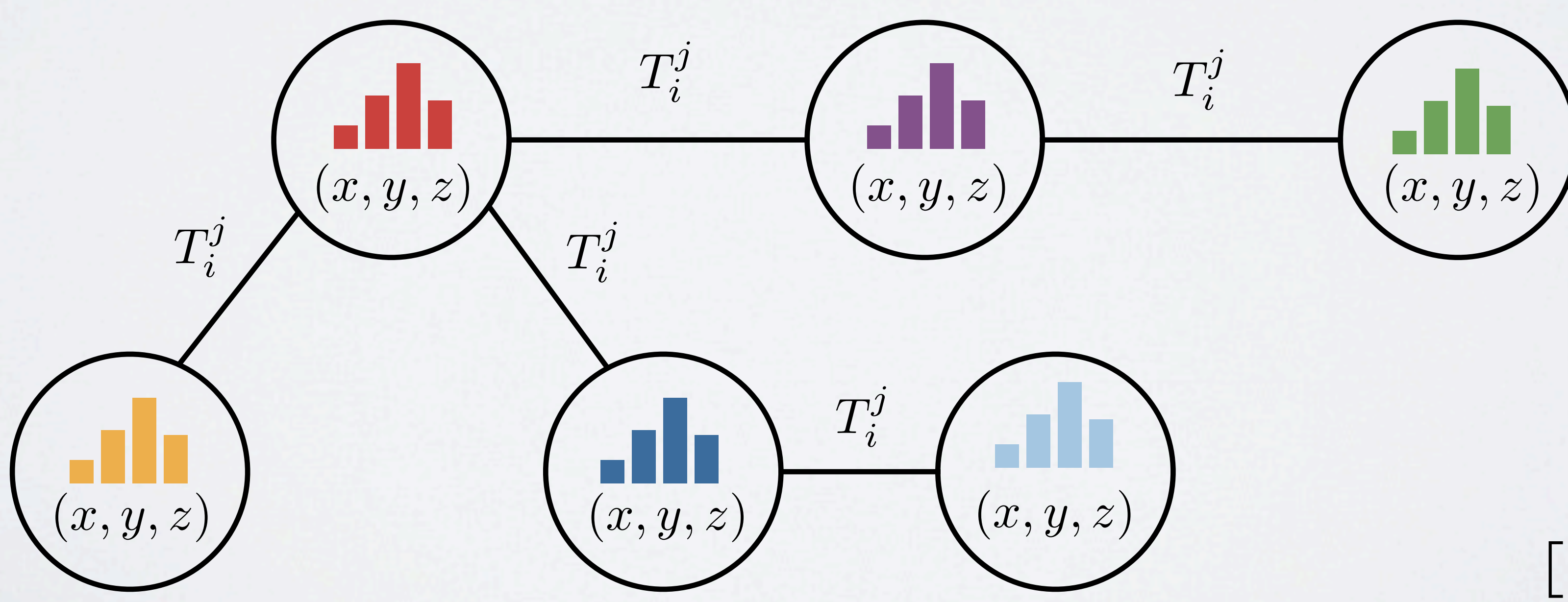
Model: Posterior over Semantic Graphs

Sensor stream Odometry Language

$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t)$$

Topology $G_t = (V_t, E_t)$ Semantic labels

Vertex poses



[RSS'13 (to appear)]

Model: Posterior over Semantic Graphs

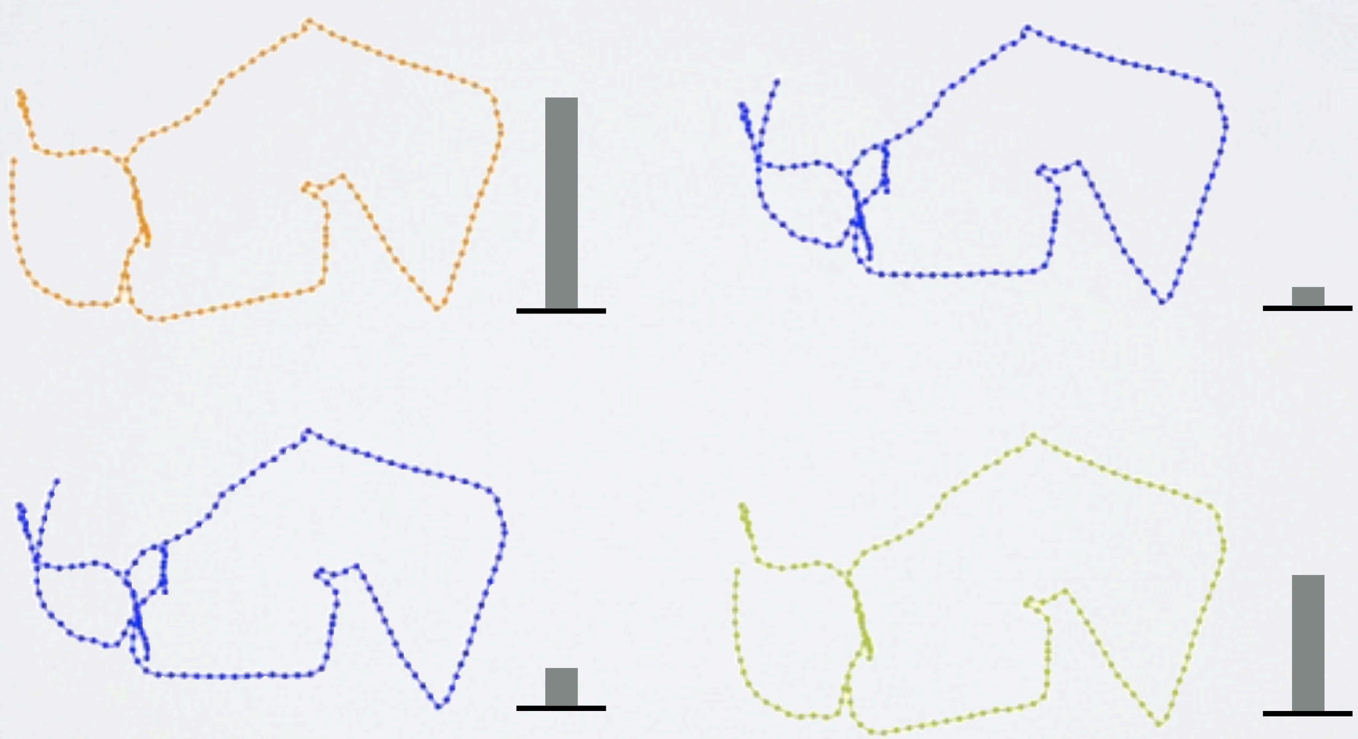
$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t) = p(L_t | X_t, G_t, z^t, u^t, \lambda^t)\ p(X_t | G_t, z^t, u^t, \lambda^t)\ p(G_t | z^t, u^t, \lambda^t)$$

[RSS'13 (to appear)]

Model: Posterior over Semantic Graphs

$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t) = p(L_t | X_t, G_t, z^t, u^t, \lambda^t)\ p(X_t | G_t, z^t, u^t, \lambda^t)\ \boxed{p(G_t | z^t, u^t, \lambda^t)}$$

Sample-based representation



[RSS'13 (to appear)]

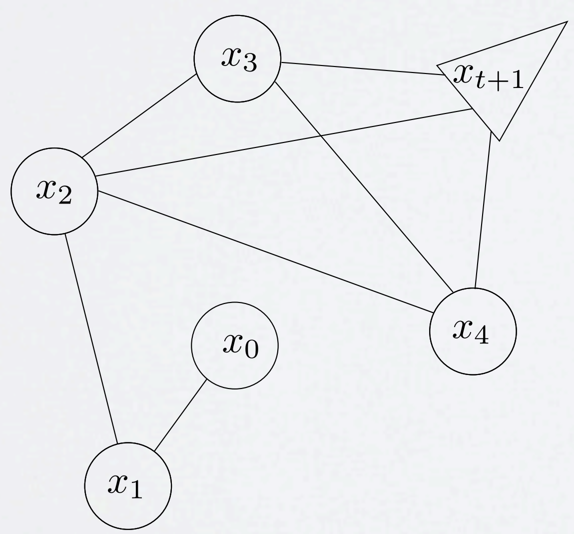
Model: Posterior over Semantic Graphs

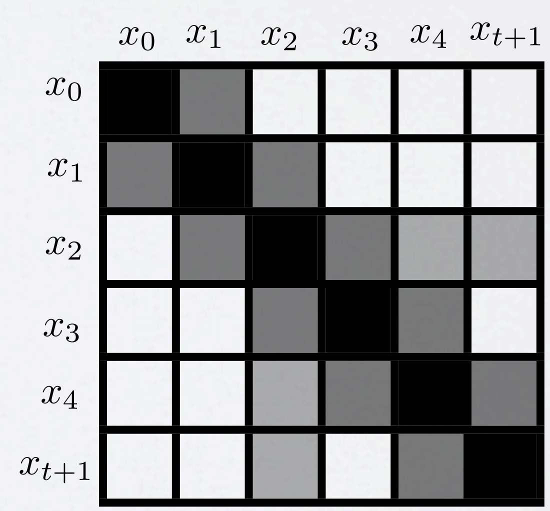
$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t) = p(L_t | X_t, G_t, z^t, u^t, \lambda^t) \boxed{p(X_t | G_t, z^t, u^t, \lambda^t)} p(G_t | z^t, u^t, \lambda^t)$$

Gaussian (information form)

Sample-based representation

$$p(X_t | G_t, z^t, u^t, \lambda^t) = \mathcal{N}^{-1}(X_t; \Sigma_t^{-1}, \eta_t)$$





[RSS'13 (to appear)]

Model: Posterior over Semantic Graphs

$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t) = \boxed{p(L_t | X_t, G_t, z^t, u^t, \lambda^t)} \; p(X_t | G_t, z^t, u^t, \lambda^t) \; p(G_t | z^t, u^t, \lambda^t)$$

Dirichlet

Gaussian (information form)

Sample-based representation

[RSS'13 (to appear)]

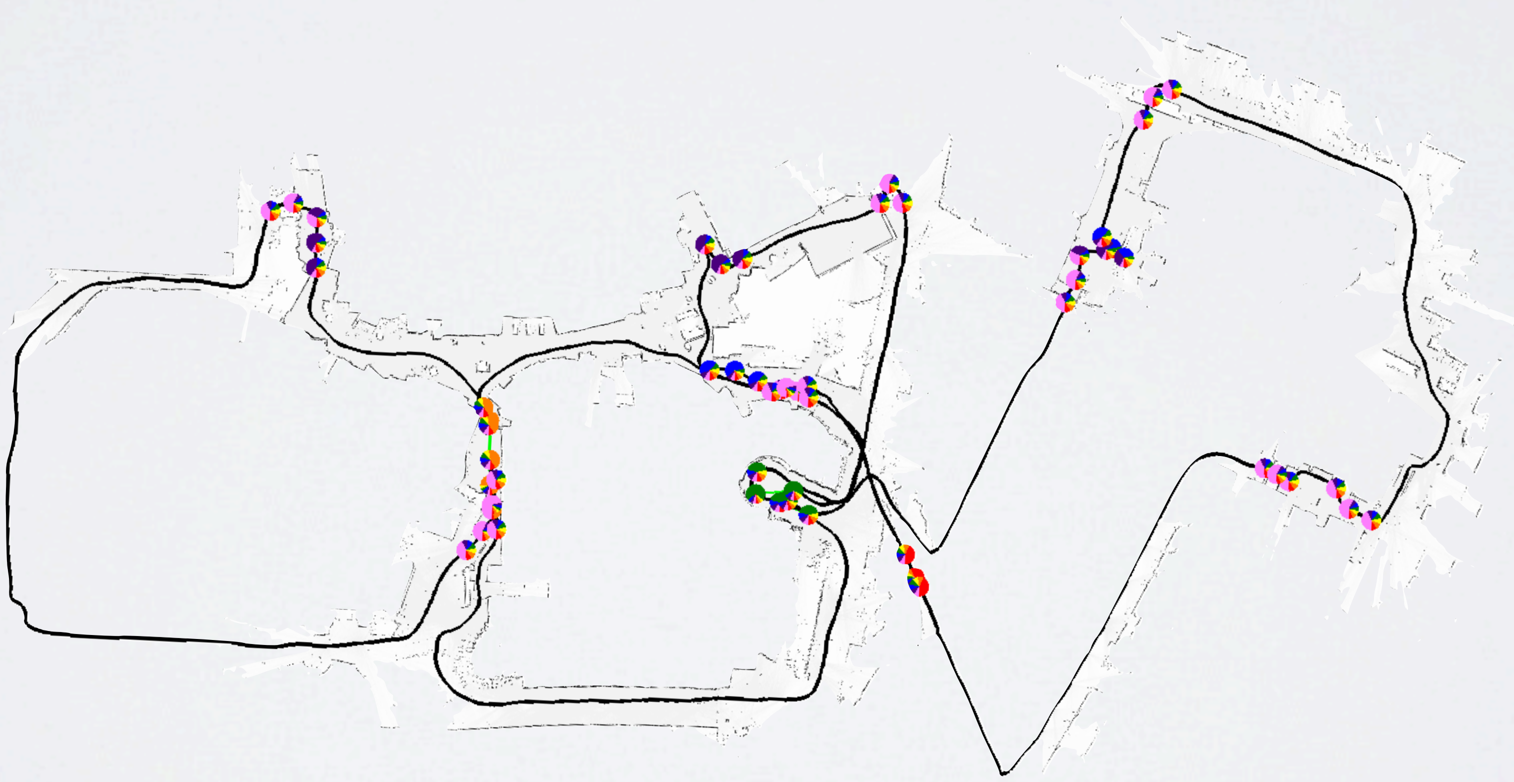
Model: Posterior over Semantic Graphs

$$p(G_t, X_t, L_t | z^t, u^t, \lambda^t) = p(L_t | X_t, G_t, z^t, u^t, \lambda^t) \; p(X_t | G_t, z^t, u^t, \lambda^t) \; p(G_t | z^t, u^t, \lambda^t)$$

Dirichlet

Gaussian
(information form)

Sample-based
representation



[RSS'13 (to appear)]

Rao-Blackwellized Particle Filter

Input: $P_{t-1} = \left\{ G_{t-1}^{(i)}, X_{t-1}^{(i)}, L_{t-1}^{(i)} w_{t-1}^{(i)} \right\} \quad (u_t, z_t, \lambda_t)$

for each particle i

1. Propose modifications to topology based on metric and semantic maps
2. Perform Bayesian update of Gaussian
3. Update Dirichlet over labels based on language
4. Update weights based on metric observations

Return: $P_t^{(i)} = \left\{ G_t^{(i)}, X_t^{(i)}, L_t^{(i)} w_t^{(i)} \right\}$

Rao-Blackwellized Particle Filter

Input: $P_{t-1} = \left\{ G_{t-1}^{(i)}, X_{t-1}^{(i)}, L_{t-1}^{(i)} w_{t-1}^{(i)} \right\} \quad (u_t, z_t, \lambda_t)$

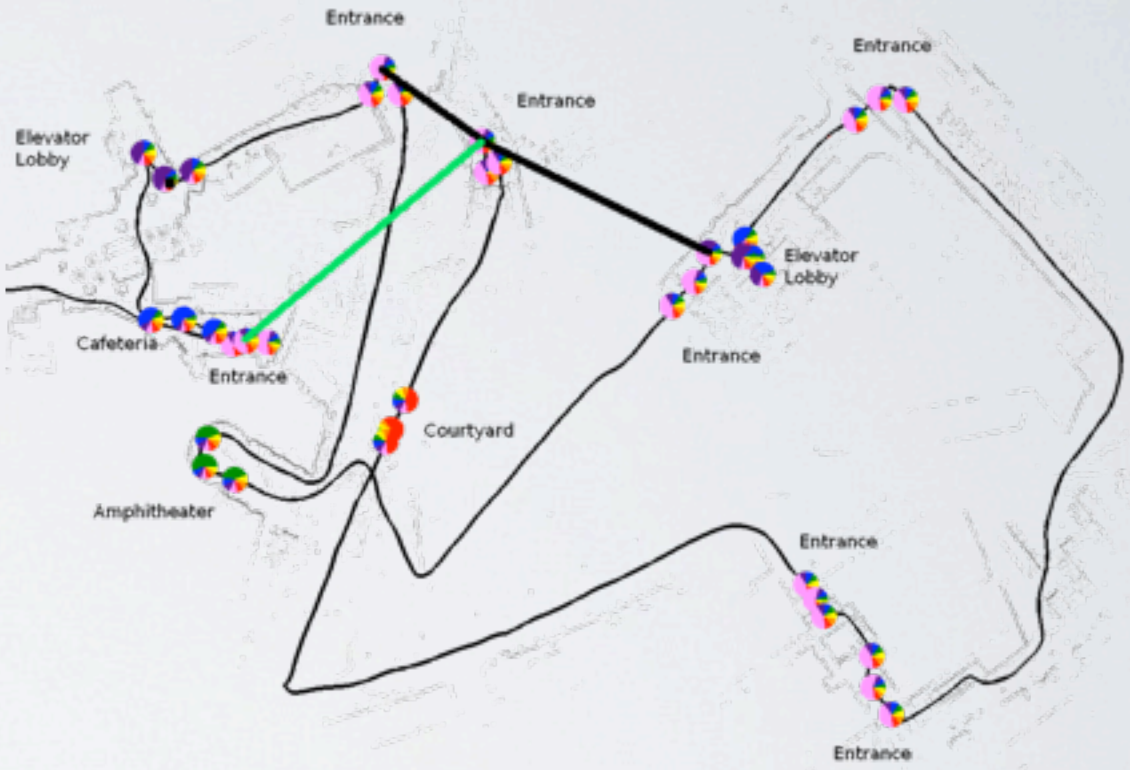
for each particle i

1 Propose modifications to topology based on metric and semantic maps

2 Perform Bayesian update of Gaussian

3 Update Dirichlet over labels based on language

4 Update weights based on metric observations



Return: $P_t^{(i)} = \left\{ G_t^{(i)}, X_t^{(i)}, L_t^{(i)} w_t^{(i)} \right\}$

Rao-Blackwellized Particle Filter

Input: $P_{t-1} = \left\{G_{t-1}^{(i)}, X_{t-1}^{(i)}, L_{t-1}^{(i)} w_{t-1}^{(i)}\right\} \quad (u_t, z_t, \lambda_t)$

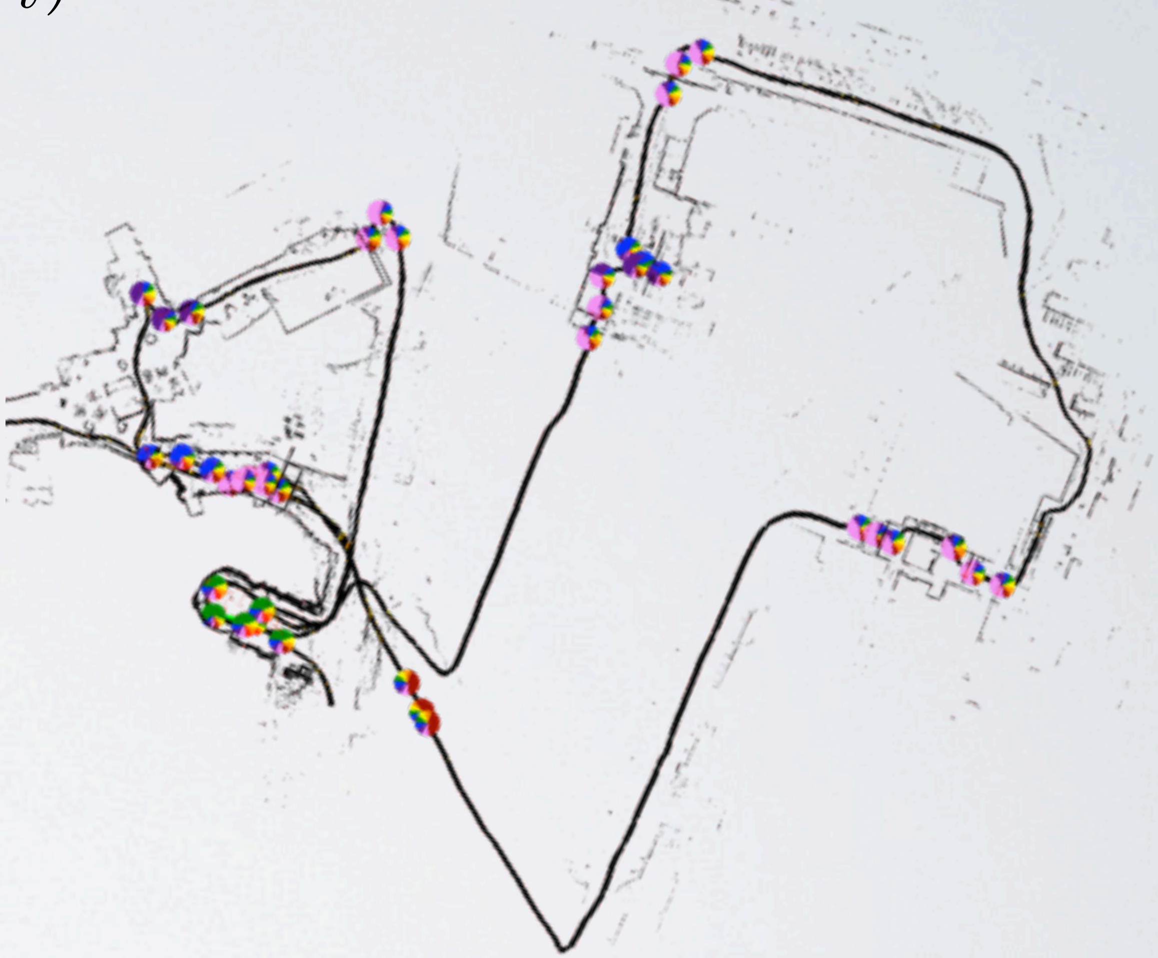
for each particle i

1 Propose modifications to topology based on metric and semantic maps

2 Perform Bayesian update of Gaussian

3 Update Dirichlet over labels based on language

4 Update weights based on metric observations



Return: $P_t^{(i)} = \left\{G_t^{(i)}, X_t^{(i)}, L_t^{(i)} w_t^{(i)}\right\}$

Rao-Blackwellized Particle Filter

Input: $P_{t-1} = \left\{ G_{t-1}^{(i)}, X_{t-1}^{(i)}, L_{t-1}^{(i)} w_{t-1}^{(i)} \right\} \quad (u_t, z_t, \lambda_t)$

for each particle i

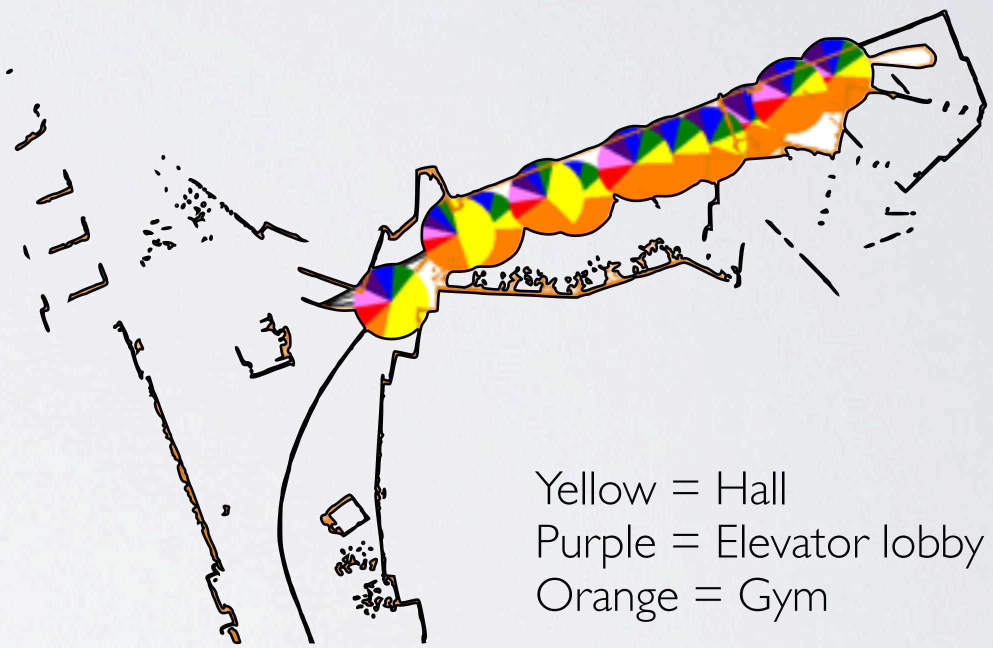
1 Propose modifications to topology based on metric and semantic maps

2 Perform Bayesian update of Gaussian

3 Update Dirichlet over labels based on language

4 Update weights based on metric observations

"the gym is down the hall"

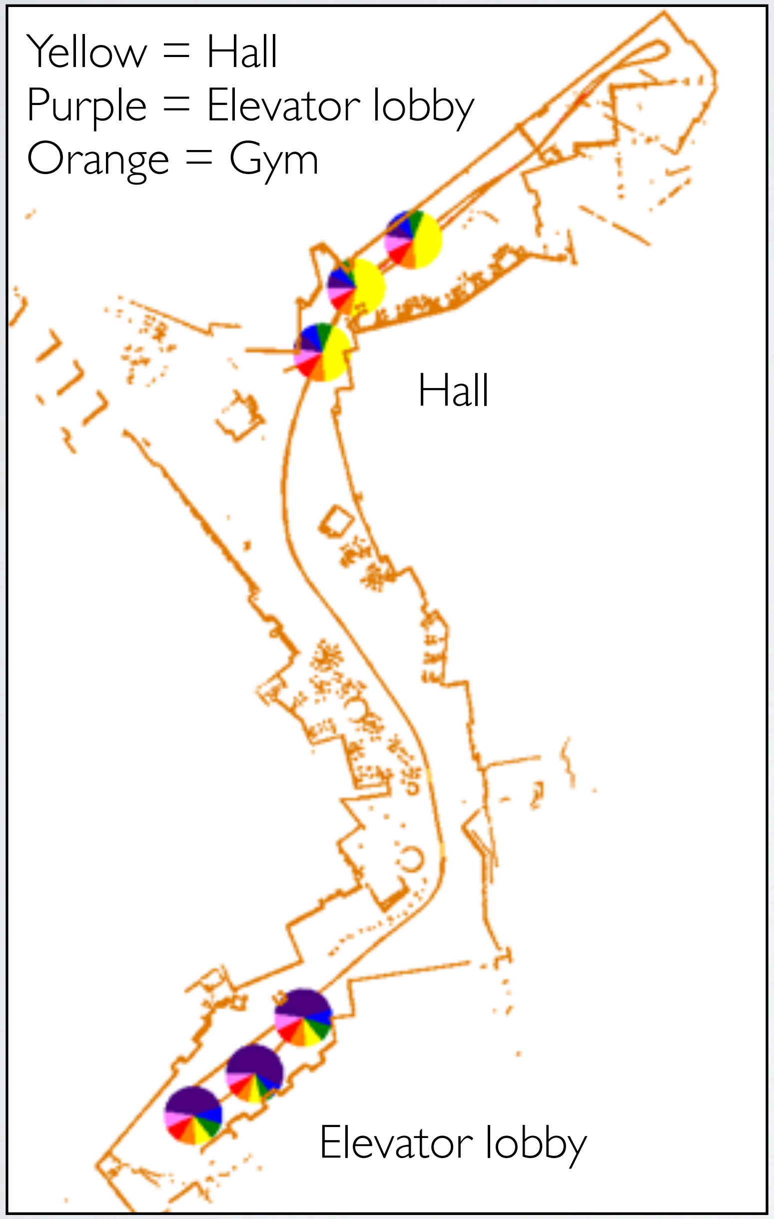


Yellow = Hall
Purple = Elevator lobby
Orange = Gym

Return: $P_t^{(i)} = \left\{ G_t^{(i)}, X_t^{(i)}, L_t^{(i)} w_t^{(i)} \right\}$

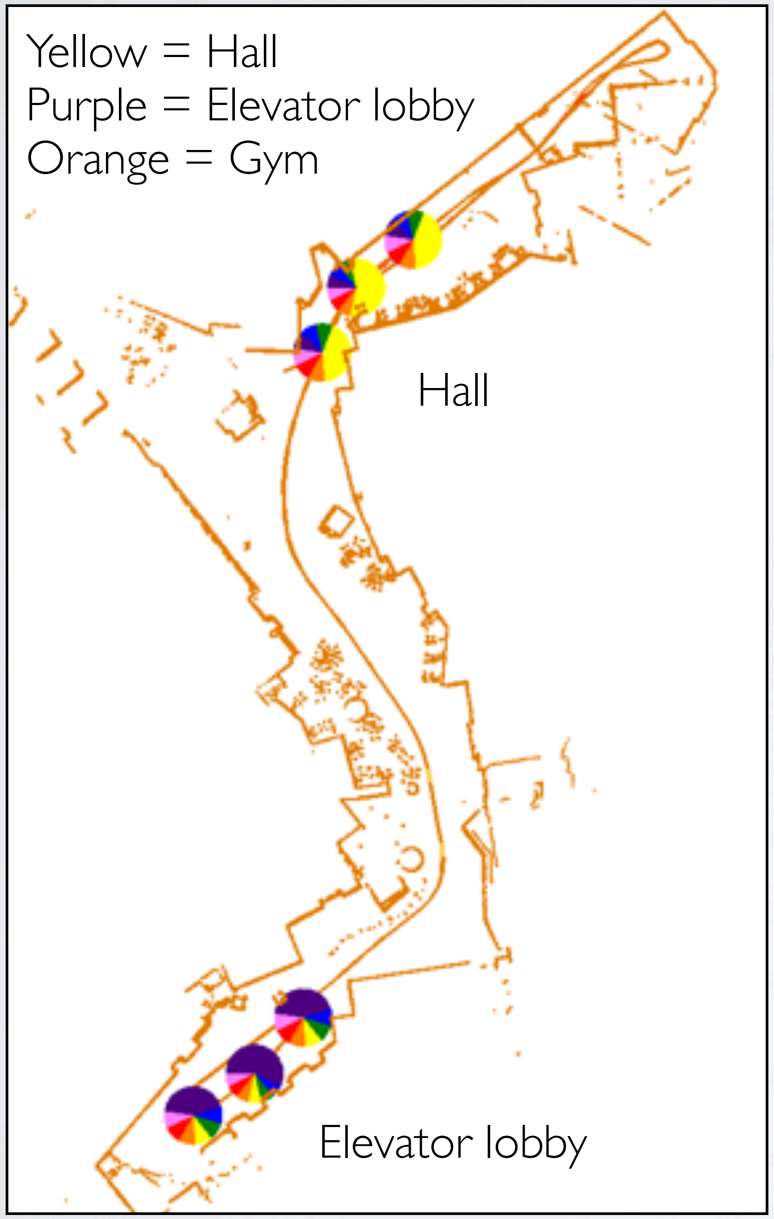
Incorporating Natural Language Descriptions

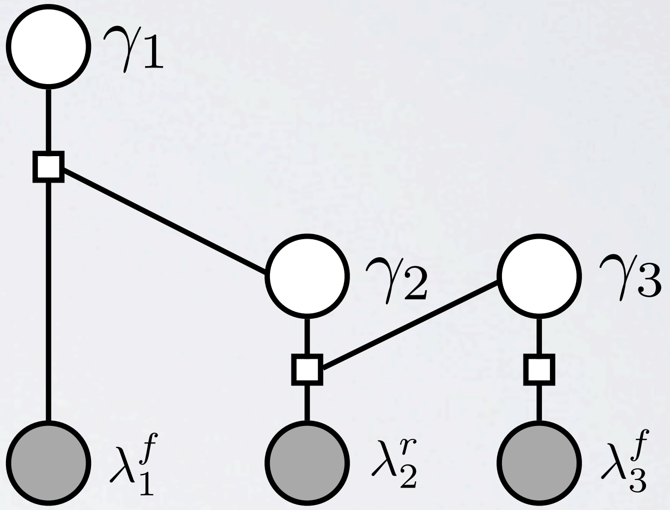
"the gym is down the hall"



Incorporating Natural Language Descriptions

"the gym is down the hall"



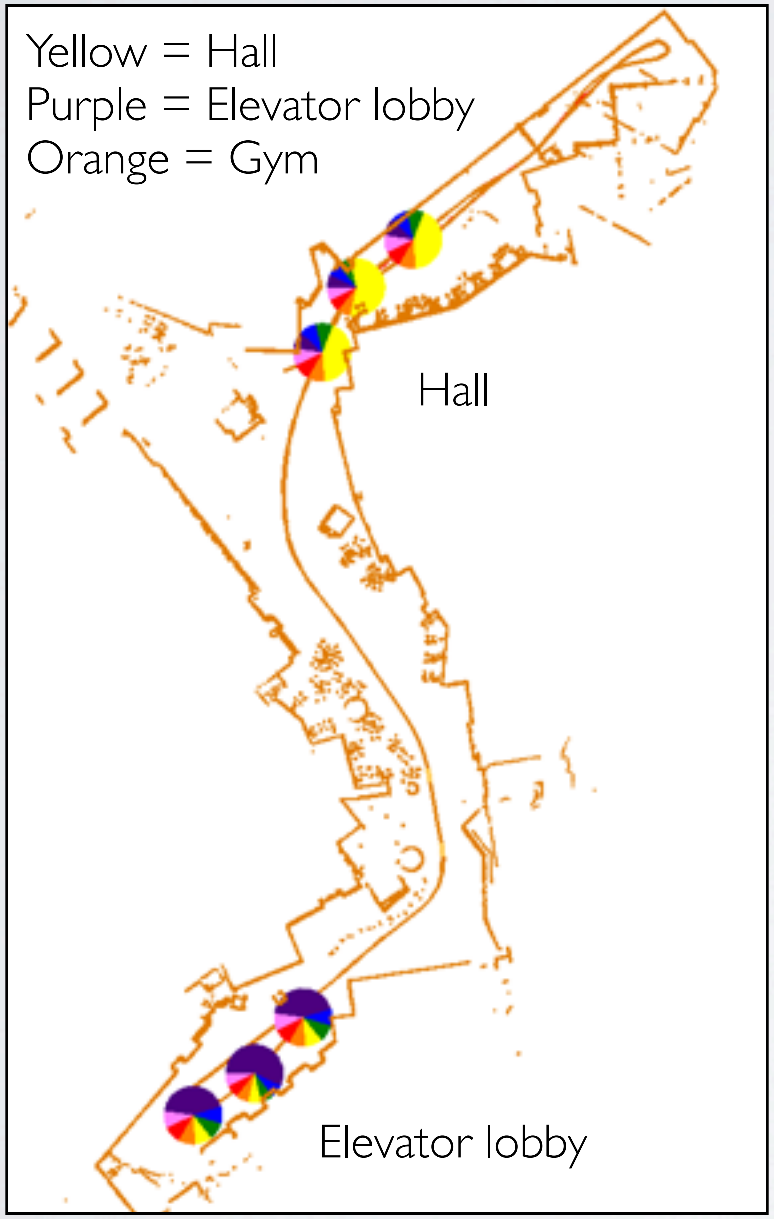


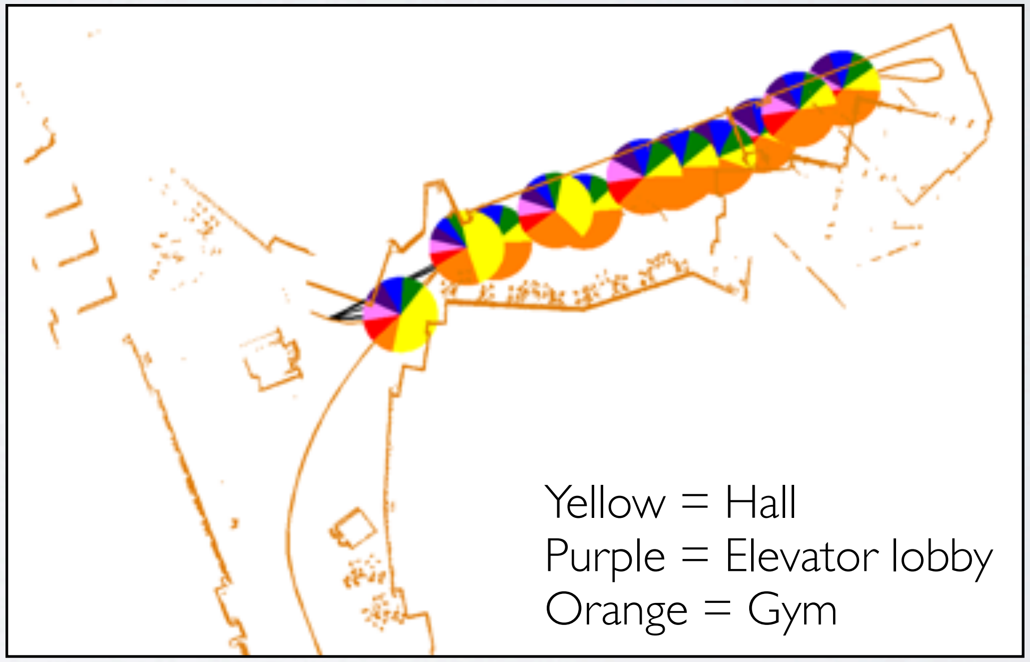
"the gym" "is down" "the hall"

$$p(L_t^{(i)}|L_{t-1}^{(i)}, G_t^{(i)}, X_t^{(i)}, \lambda_t) = \sum_{\gamma} p(L_t^{(i)}|\gamma, L_{t-1}^{(i)}, \lambda_t) \times p(\gamma|L_{t-1}^{(i)}, G_t^{(i)}, X_t^{(i)}, \lambda_t)$$

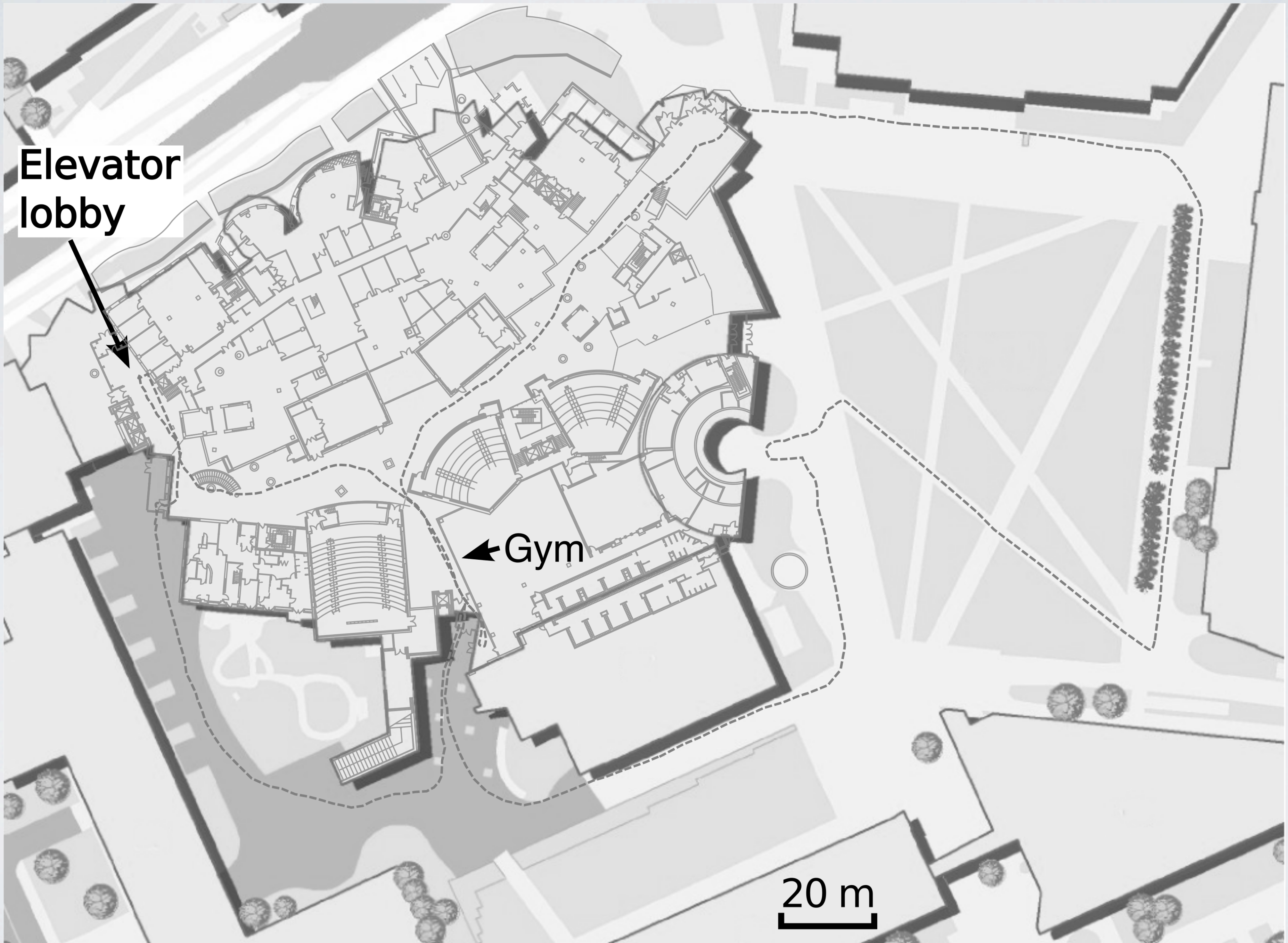
Incorporating Natural Language Descriptions

"the gym is down the hall"

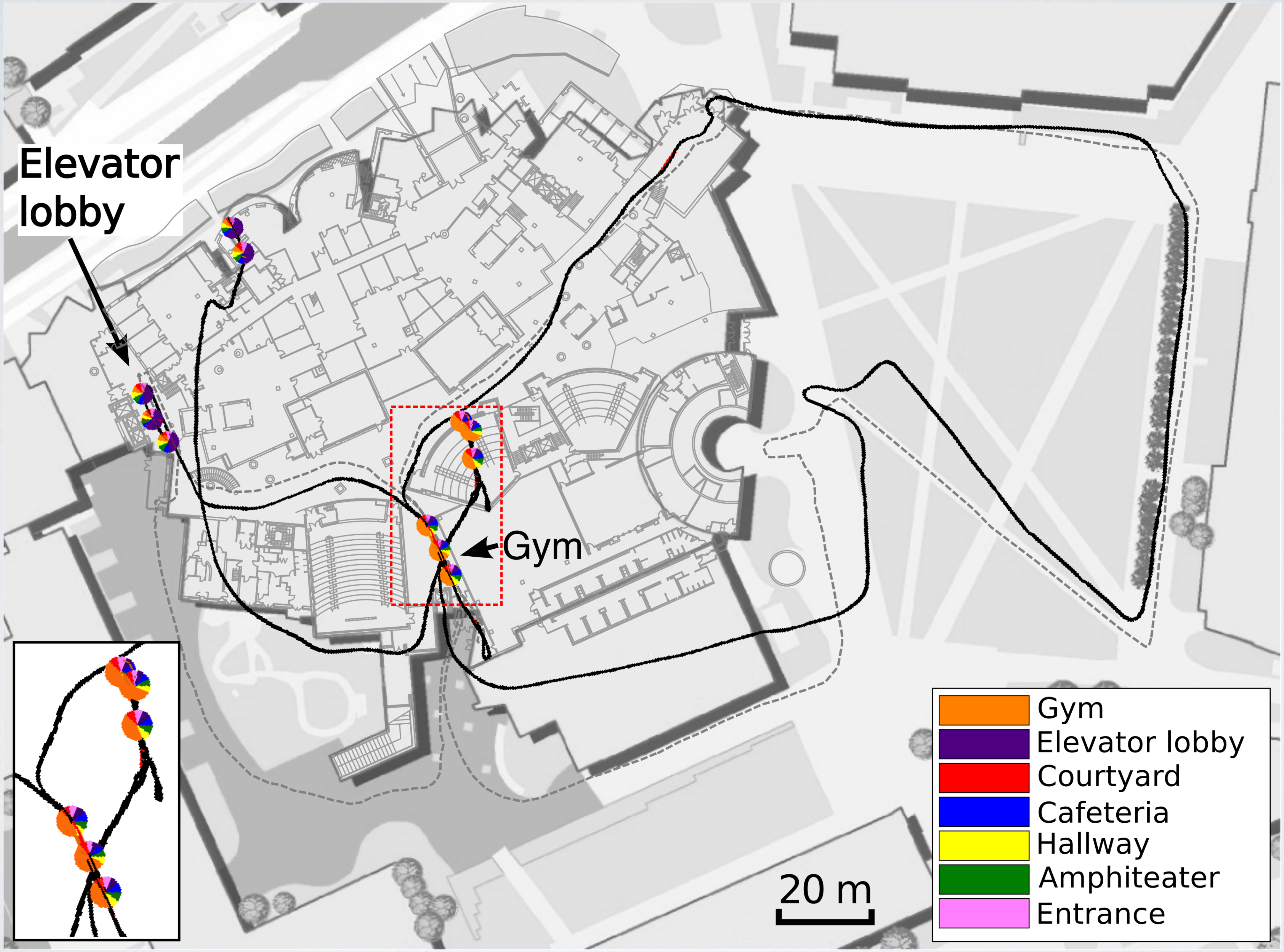




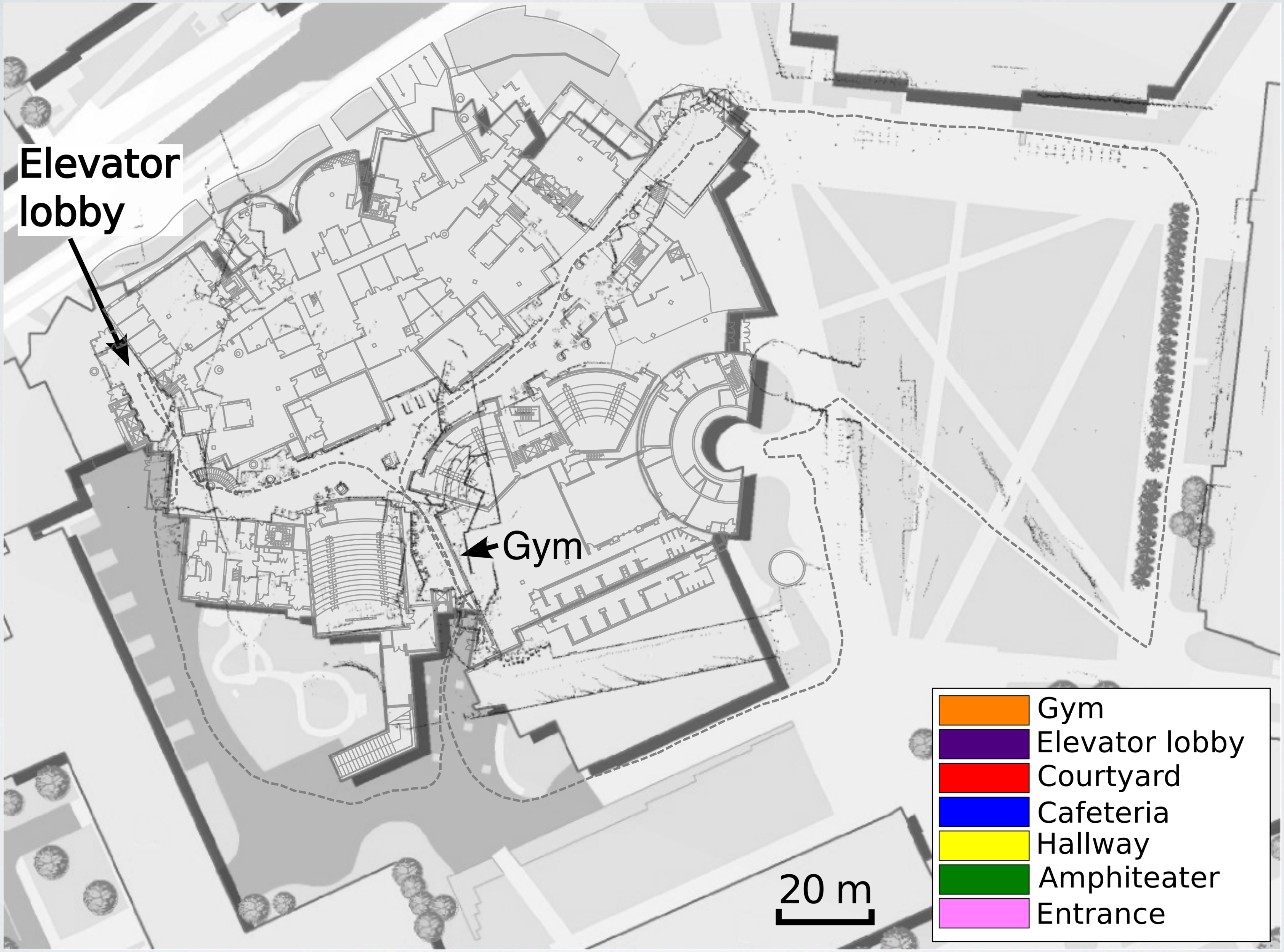
Maps without Language Constraints



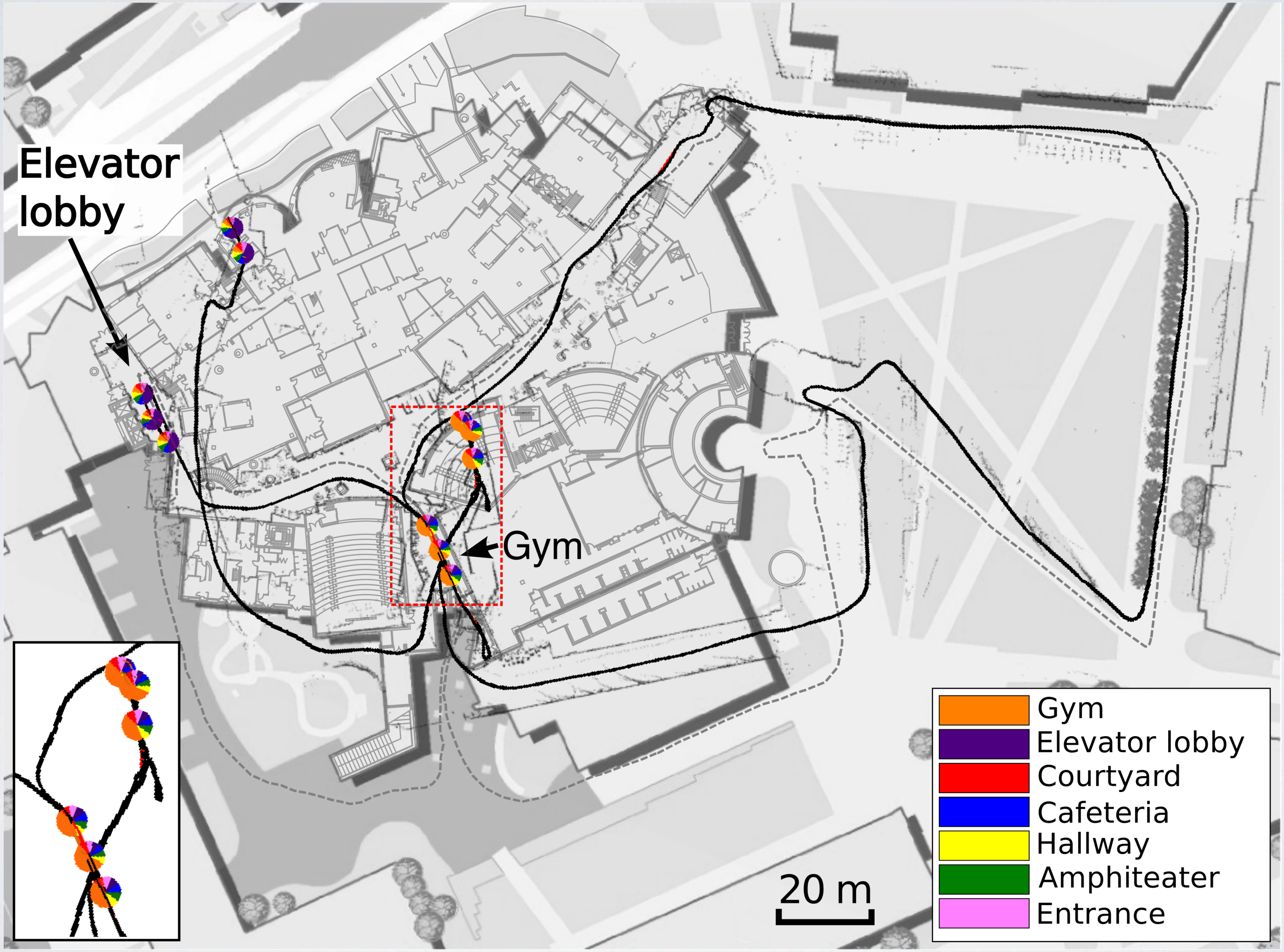
Maps without Language Constraints



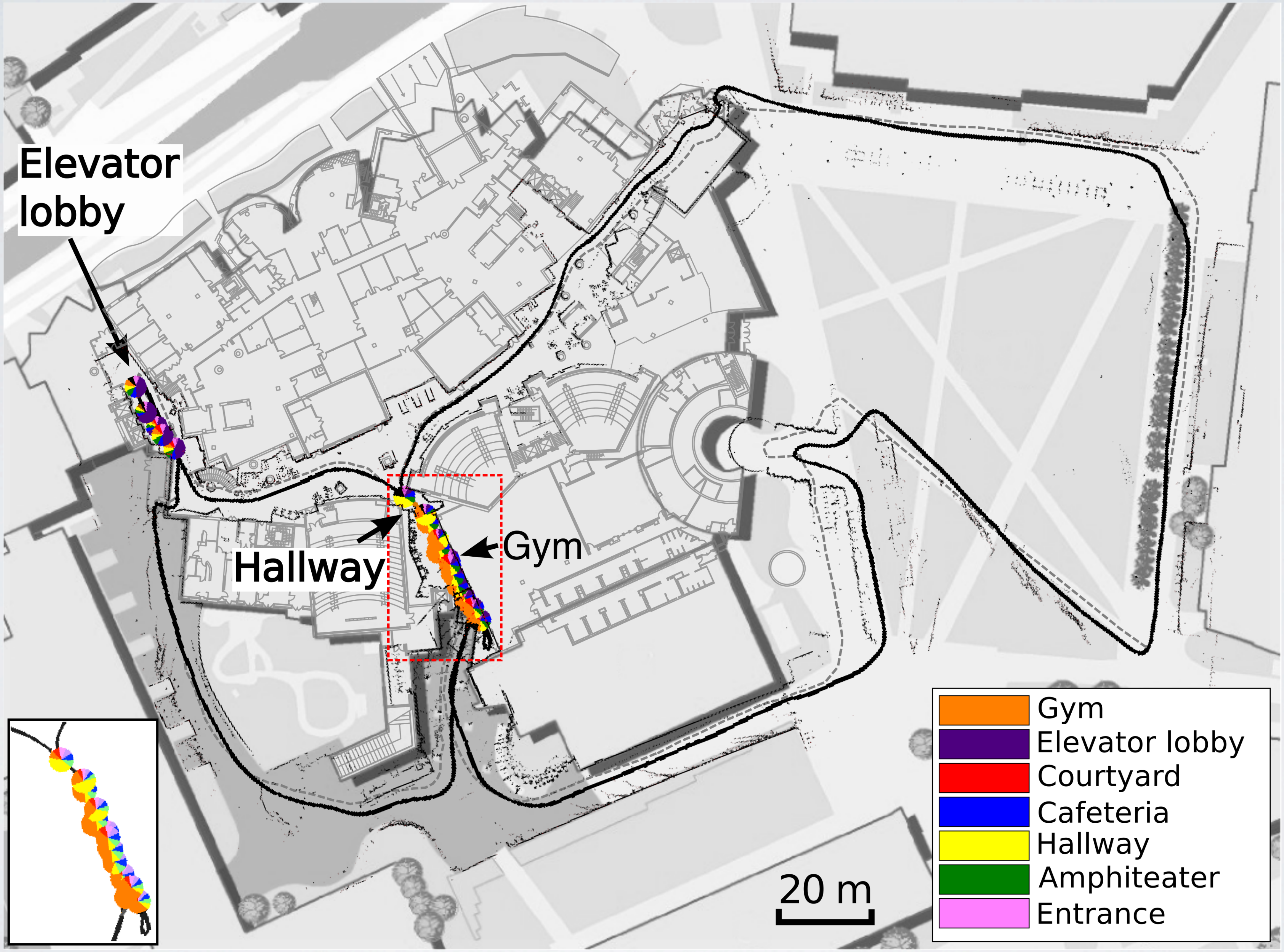
Maps without Language Constraints



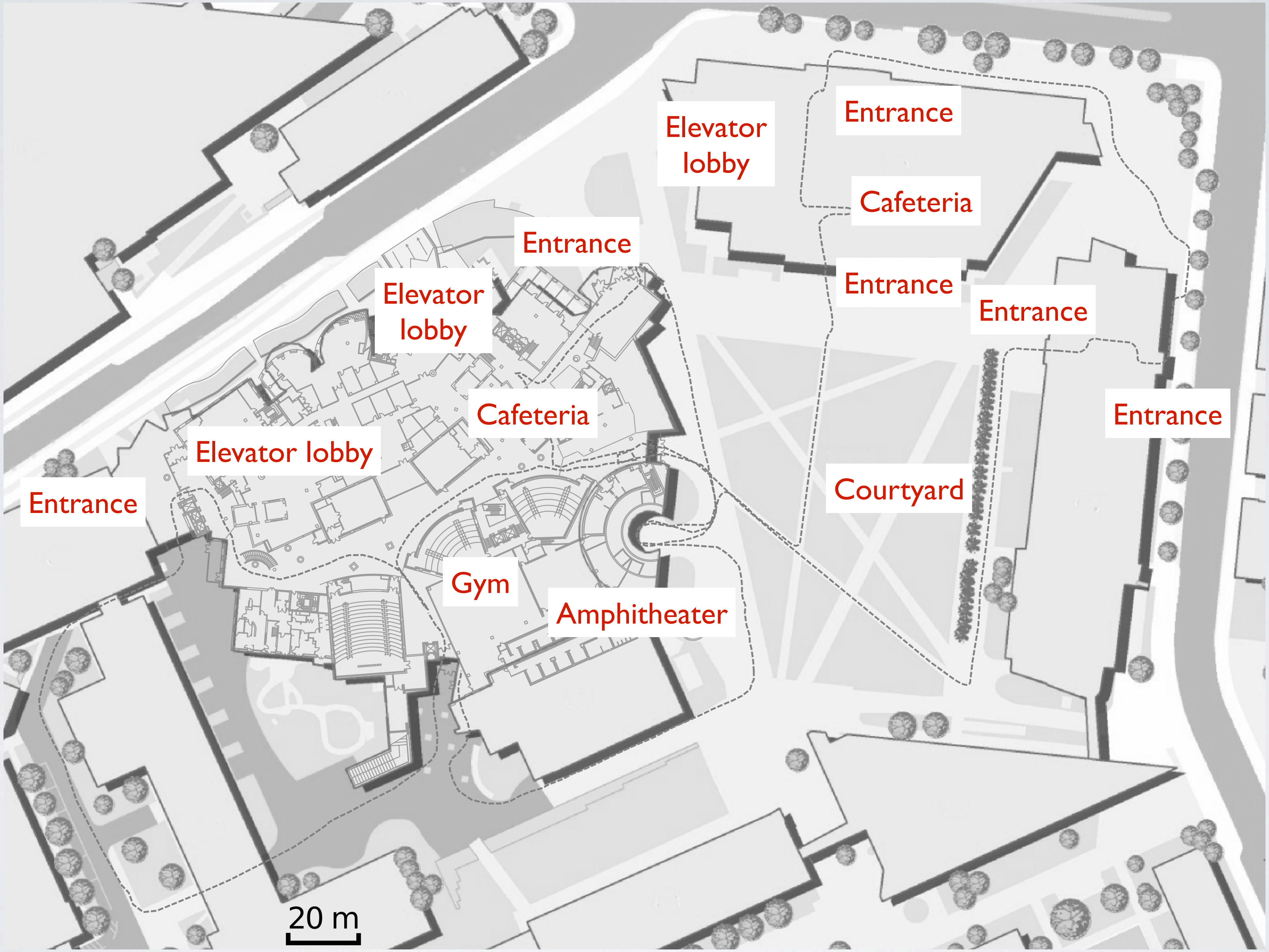
Maps without Language Constraints



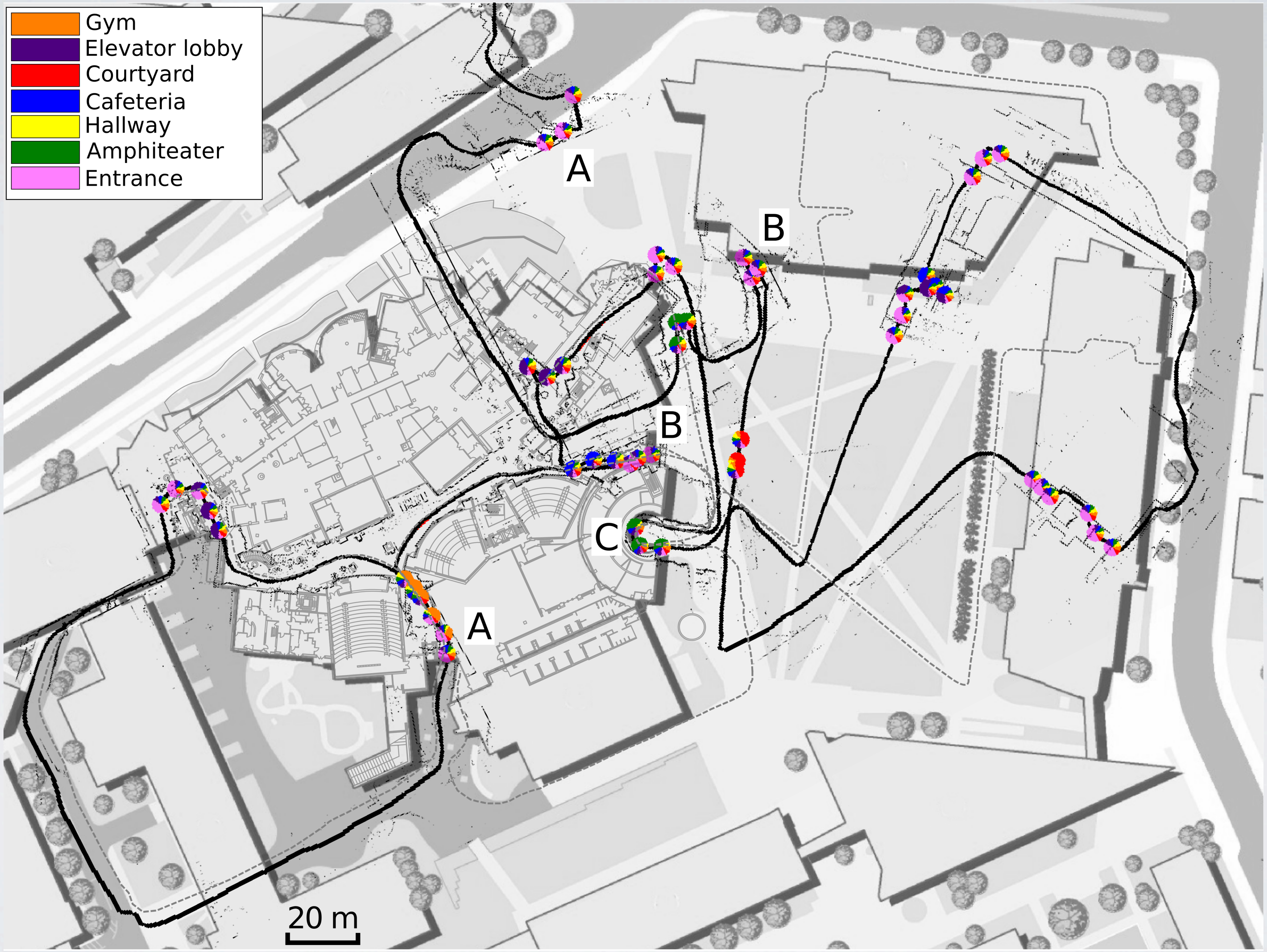
Maps with Language Constraints



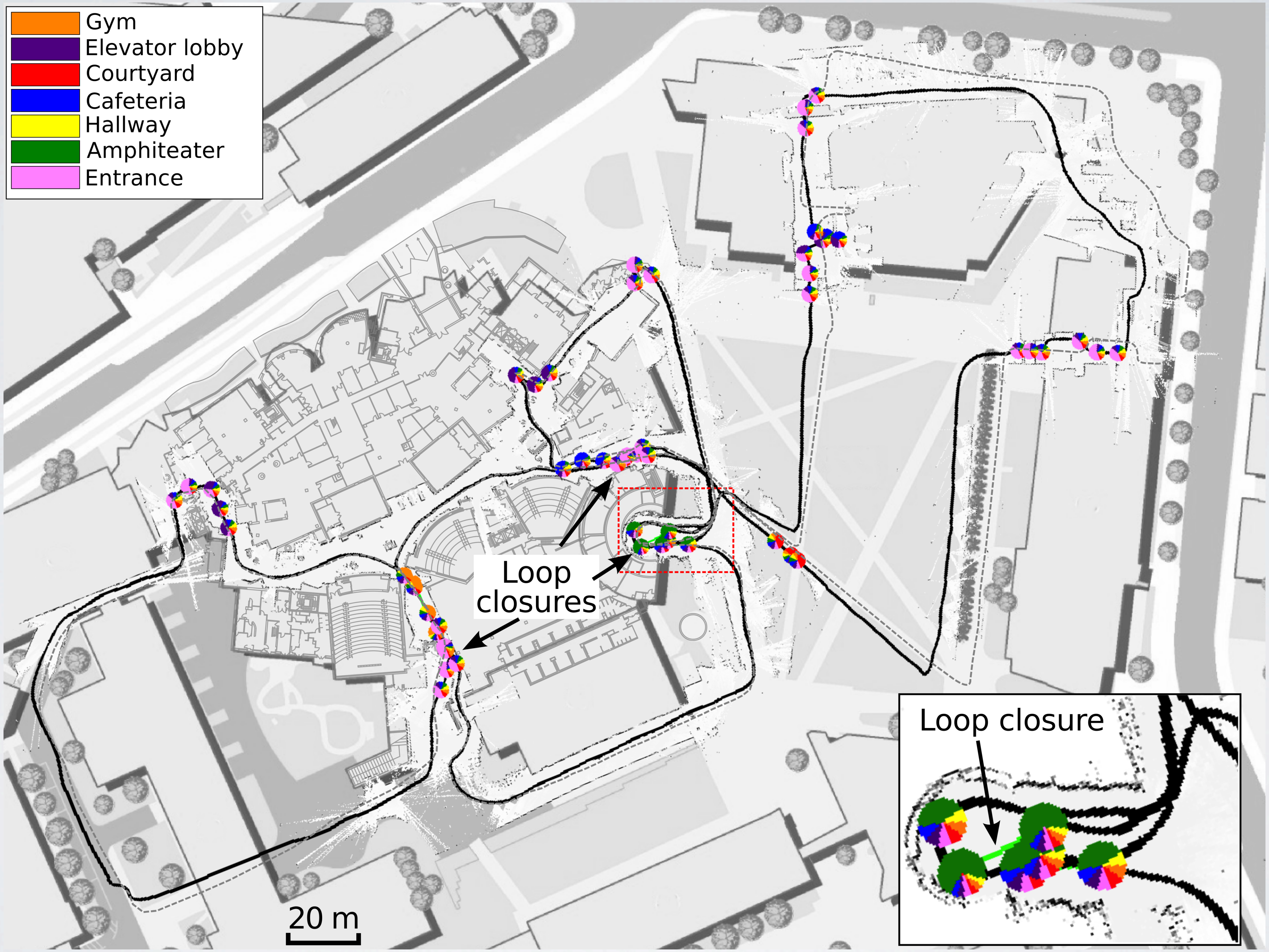
Maps without Language Constraints



Maps without Language Constraints



Maps without Language Constraints



Autonomous Narrated Tour

Autonomous Narrated Tour

Questions?