Discriminative Label Propagation for Multi-object Tracking with Sporadic Appearance Features

Amit Kumar K.C.
Christophe De Vleeschouwer

Organization

- Context/Introduction
- Graph construction
- Label propagation framework
- Optimization
- Results

Pipeline

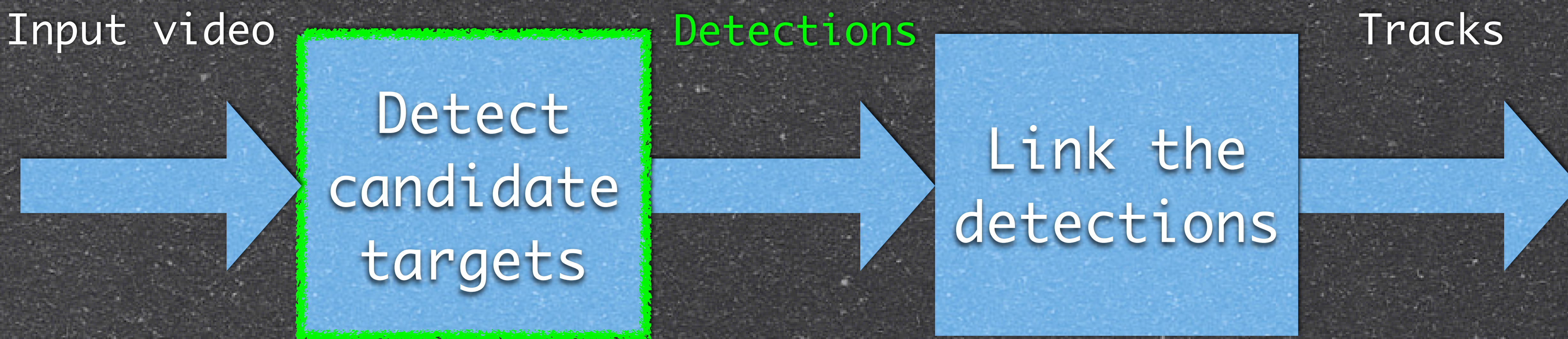

Input video
Detect candidate targets
Detections
Link the detections
Tracks

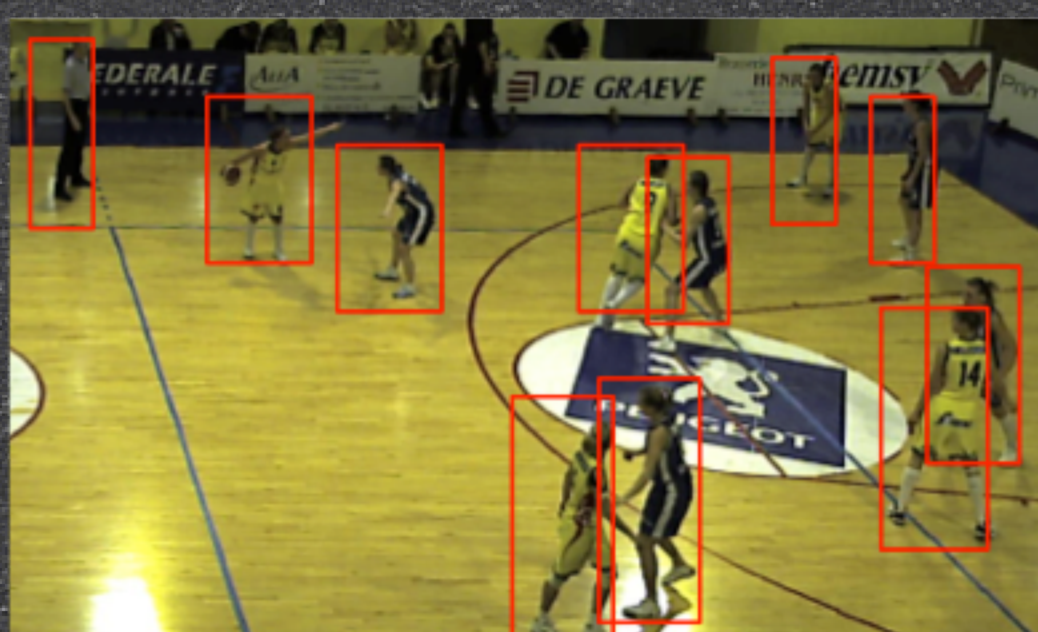
3

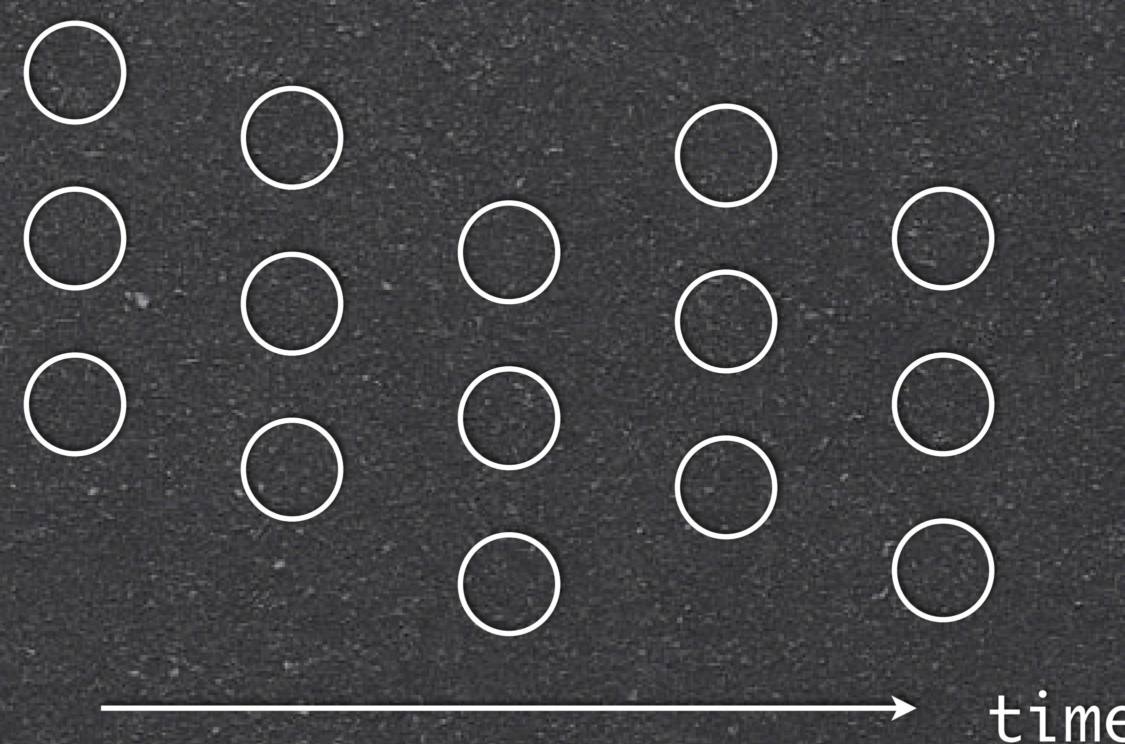
Pipeline

Input video → **Detect candidate targets** → Detections → **Link the detections** → Tracks

Pipeline


Input video
Detect candidate targets
Detections
Link the detections
Tracks




time

Pipeline

Input video

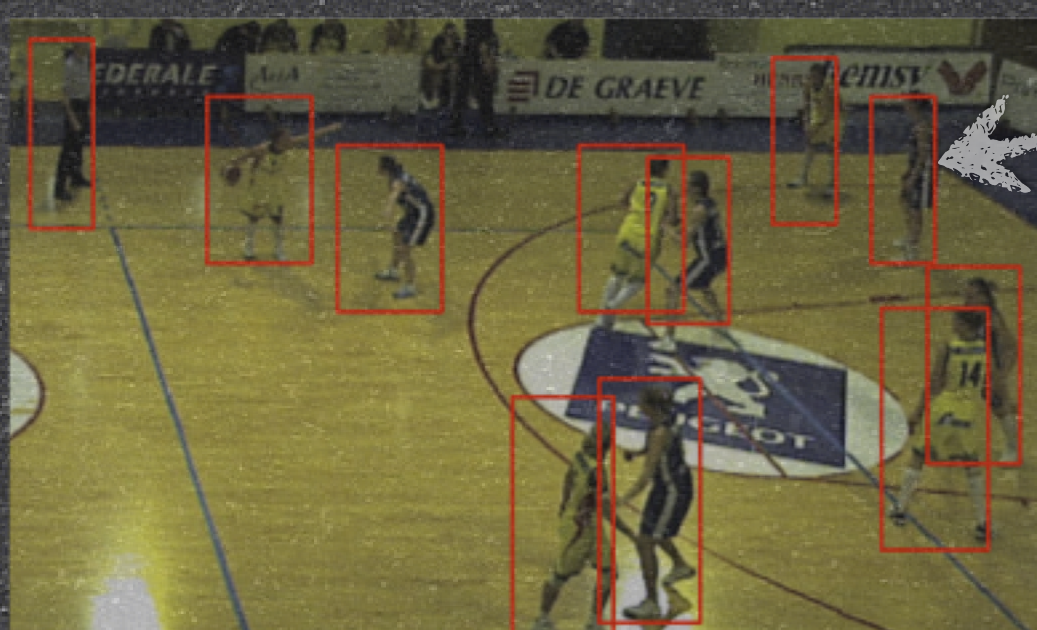
Detect candidate targets

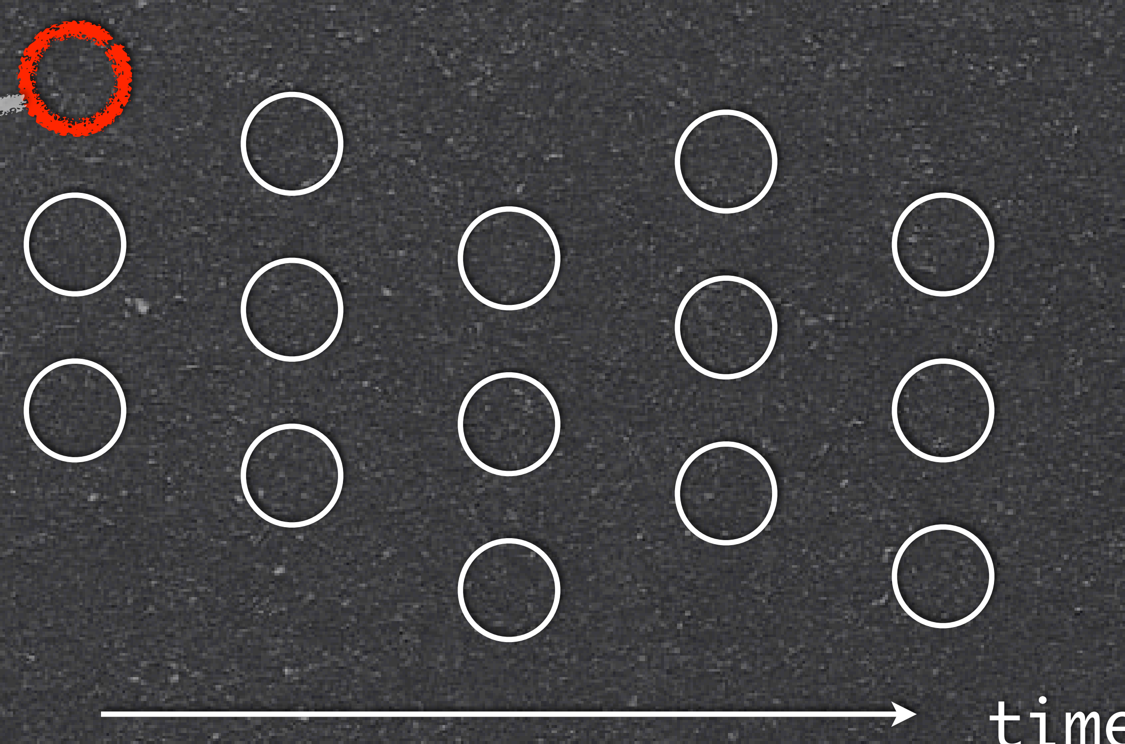
Det

Node

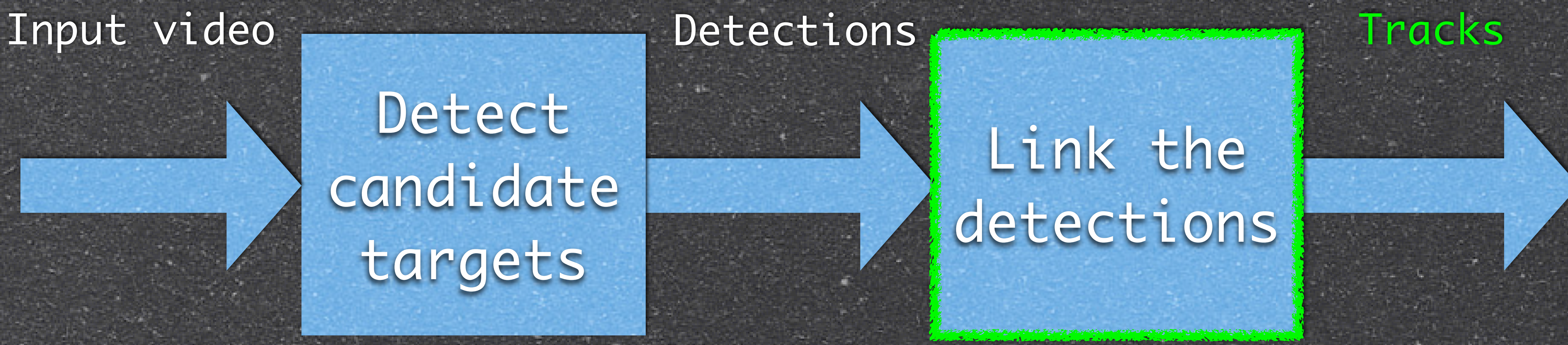
A hypothetical variable

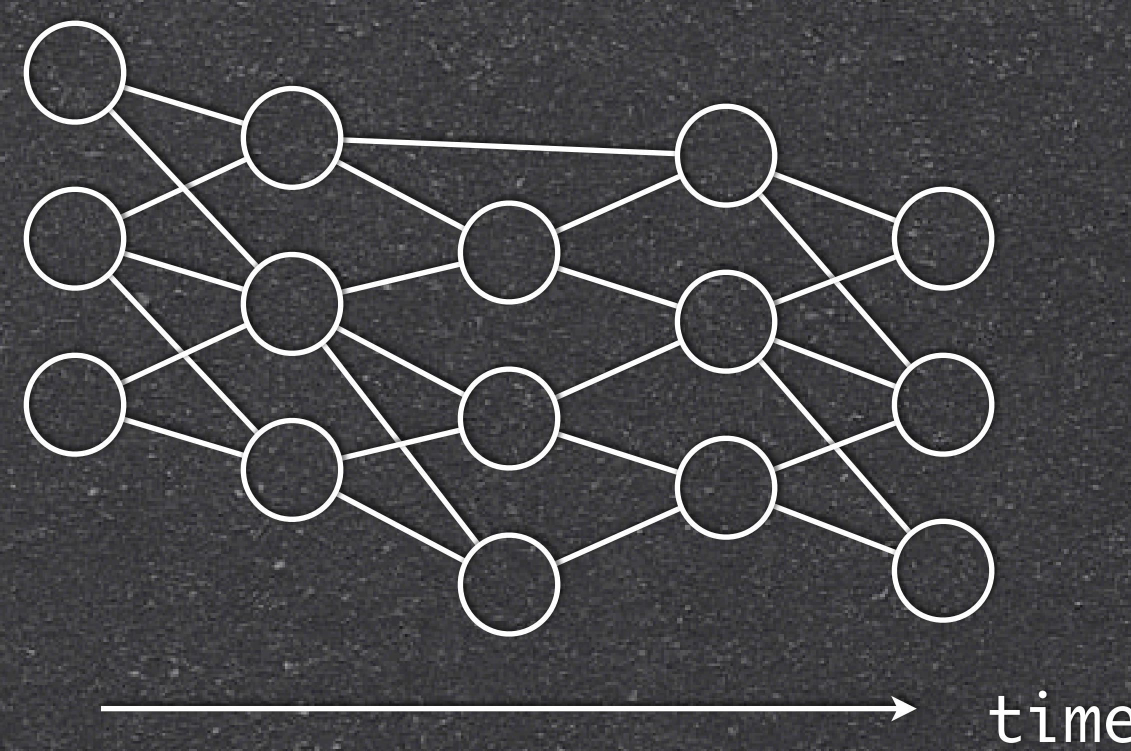
- Time instant
- Position
- Bounding box
- Color histogram
- and many more...



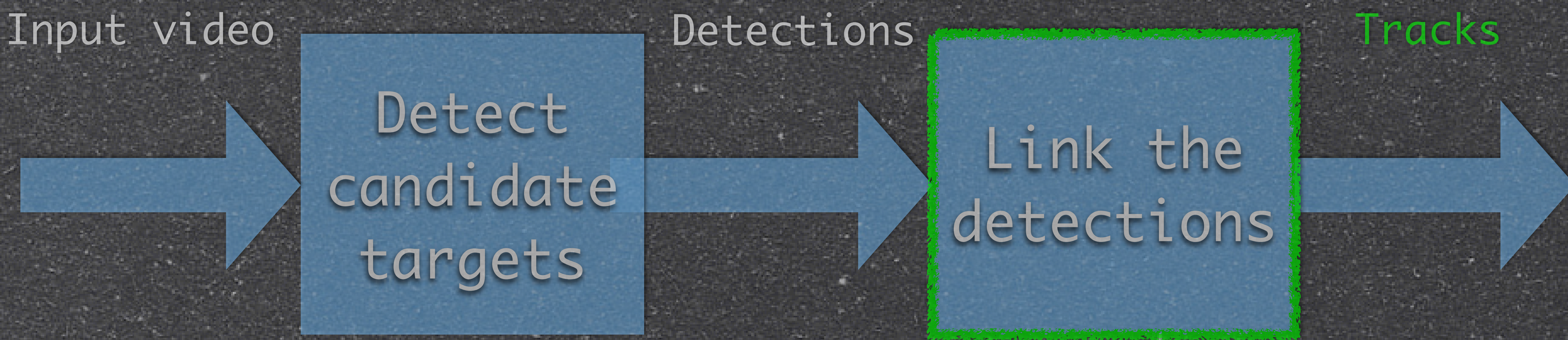


Pipeline


Input video
Detect candidate targets
Detections
Link the detections
Tracks

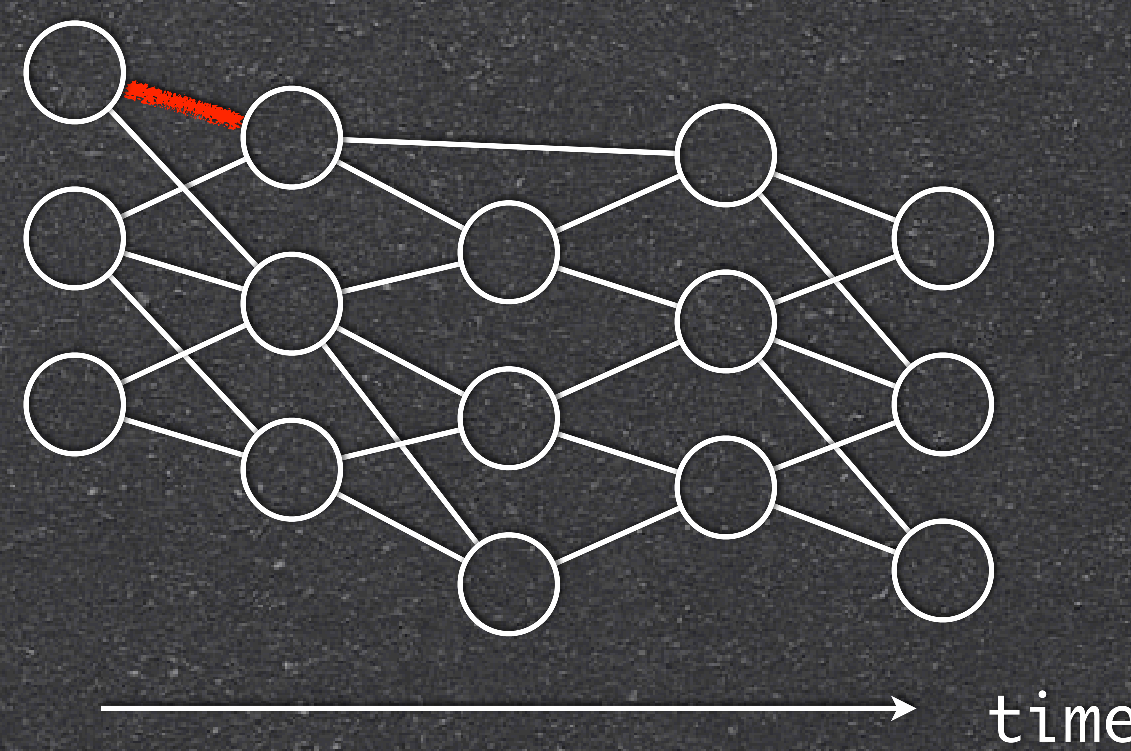

time

Pipeline

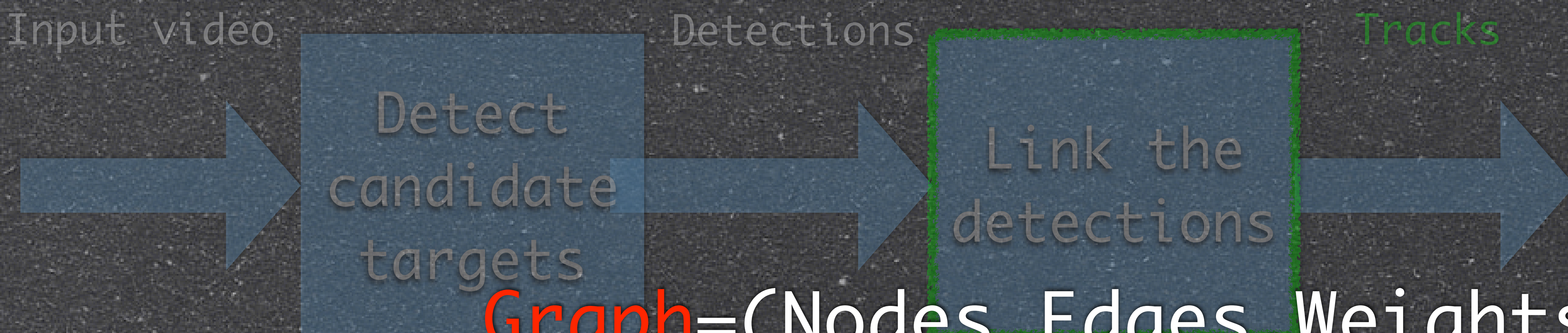


Edge

- Relationship between two nodes
- Assigned some <u>weight</u>



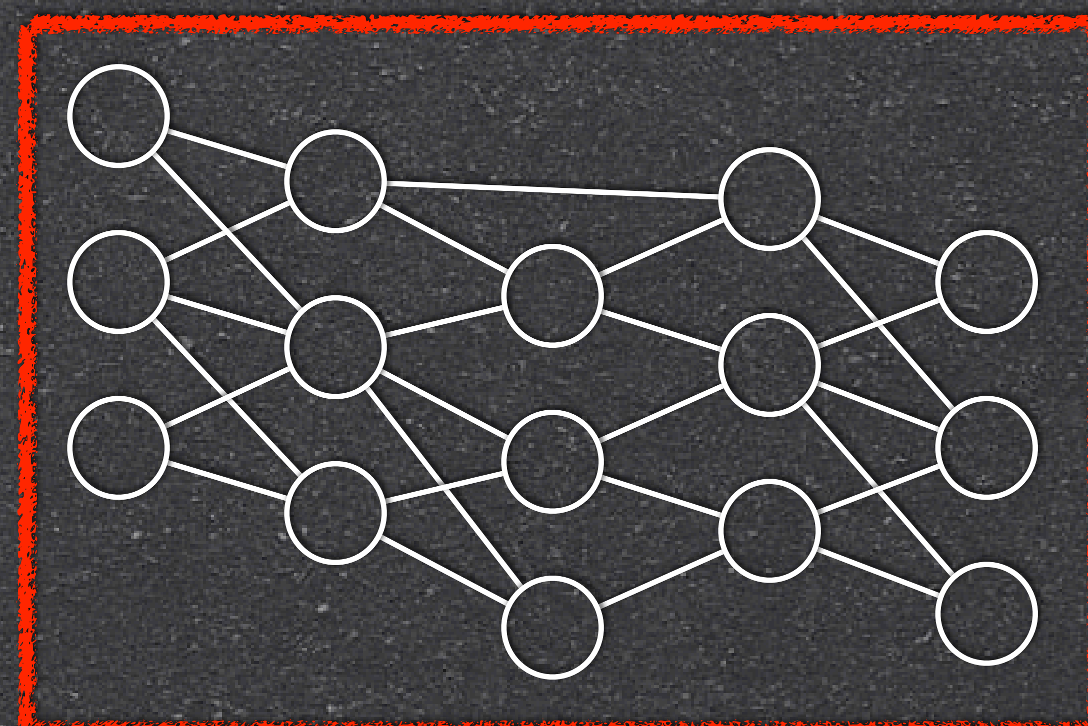
Pipeline



Graph=(Nodes,Edges,Weights)

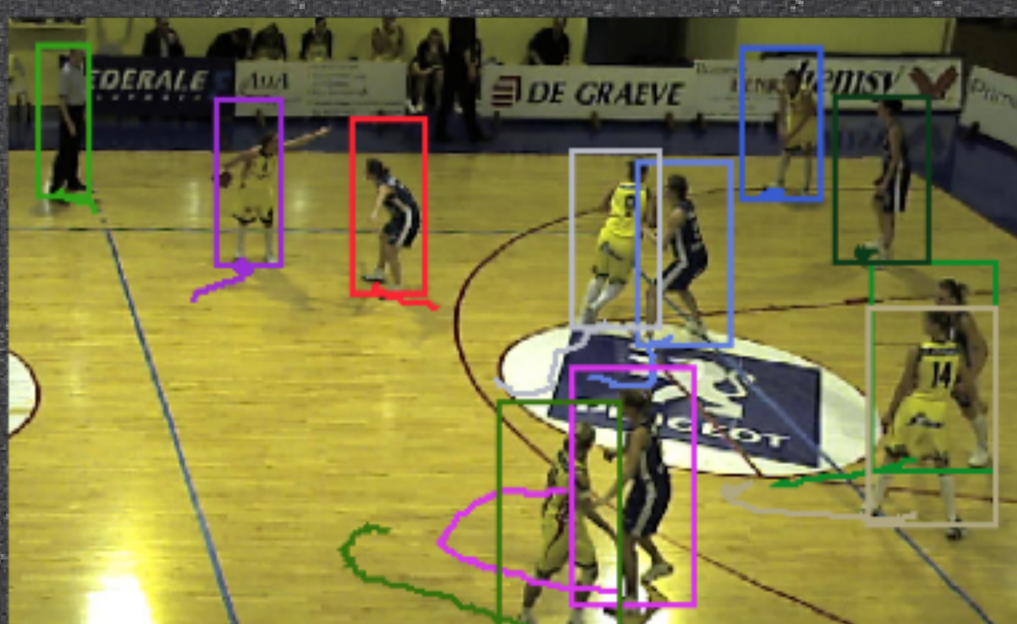
Edge

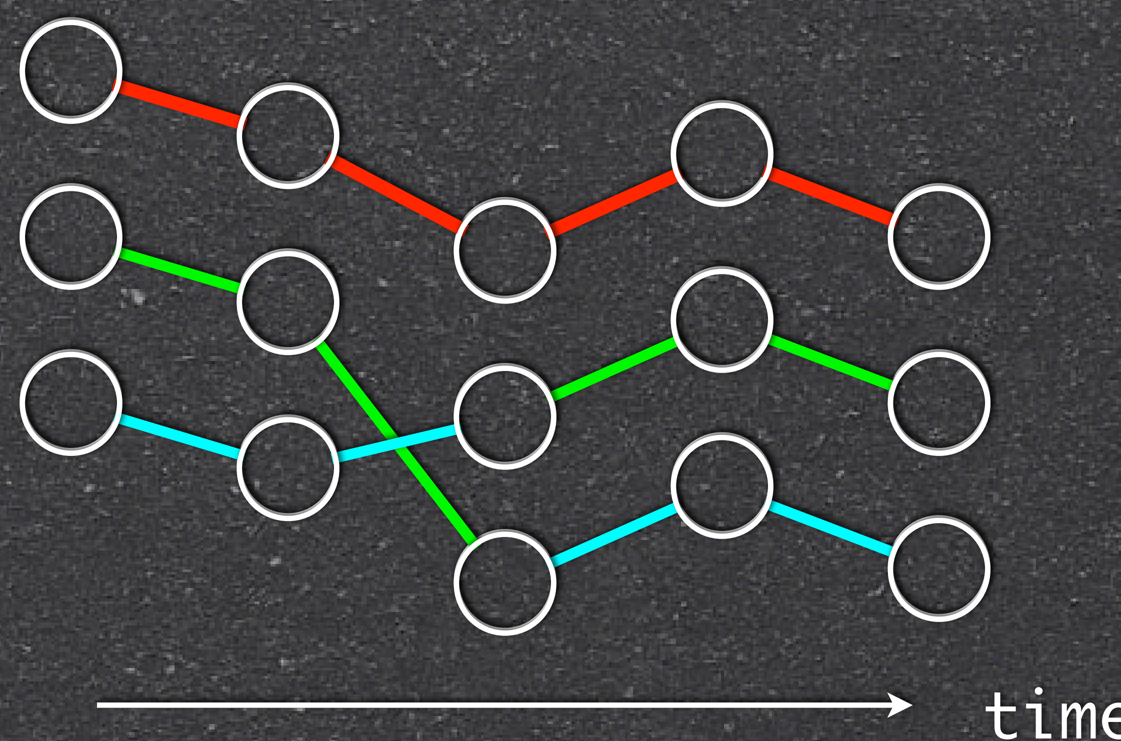
- Relationship between two nodes
- Assigned some weight



Pipeline

Input video → **Detect candidate targets** → Detections → **Link the detections** → Tracks





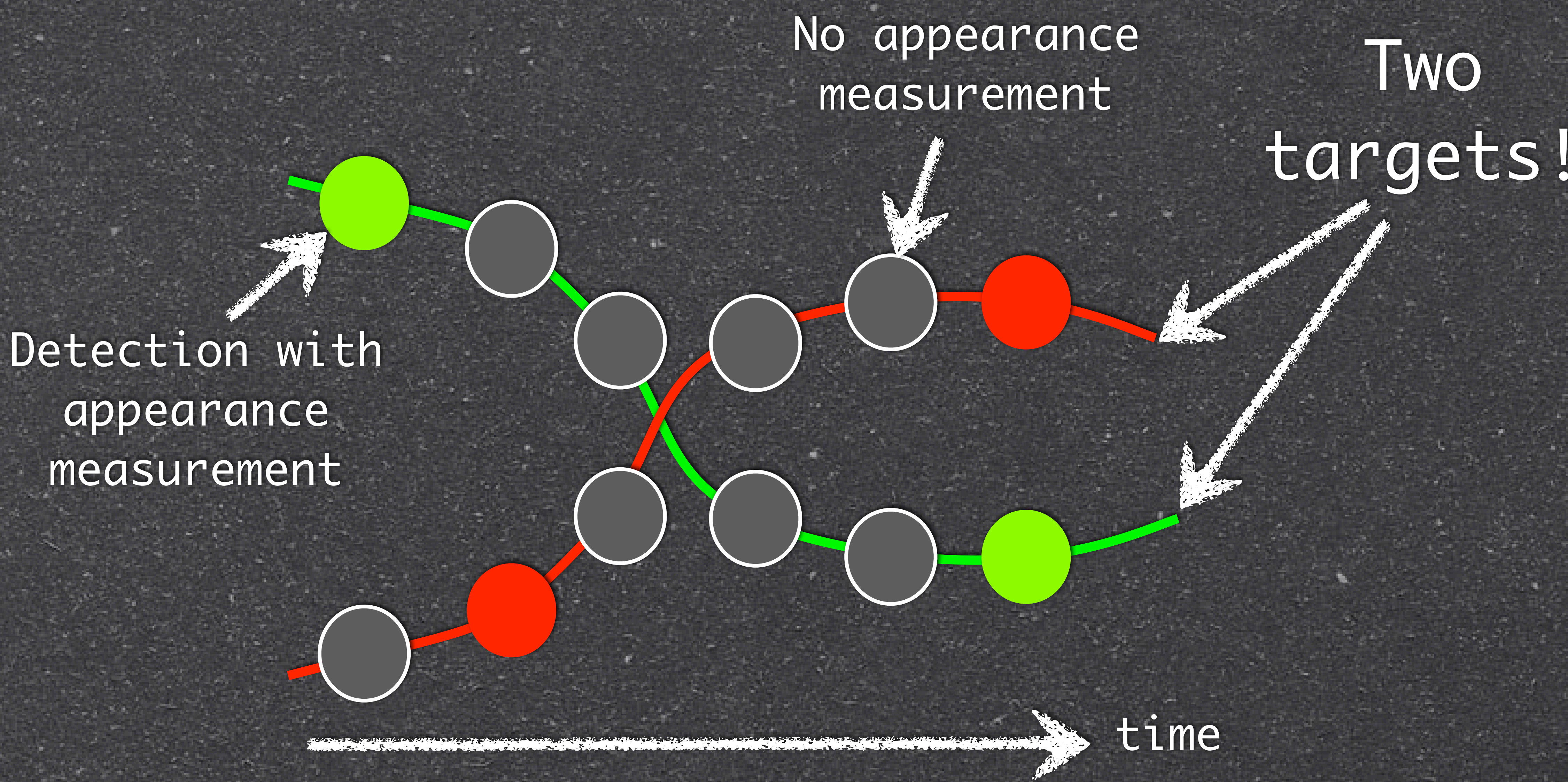
10

Pipeline

Focus of the talk!

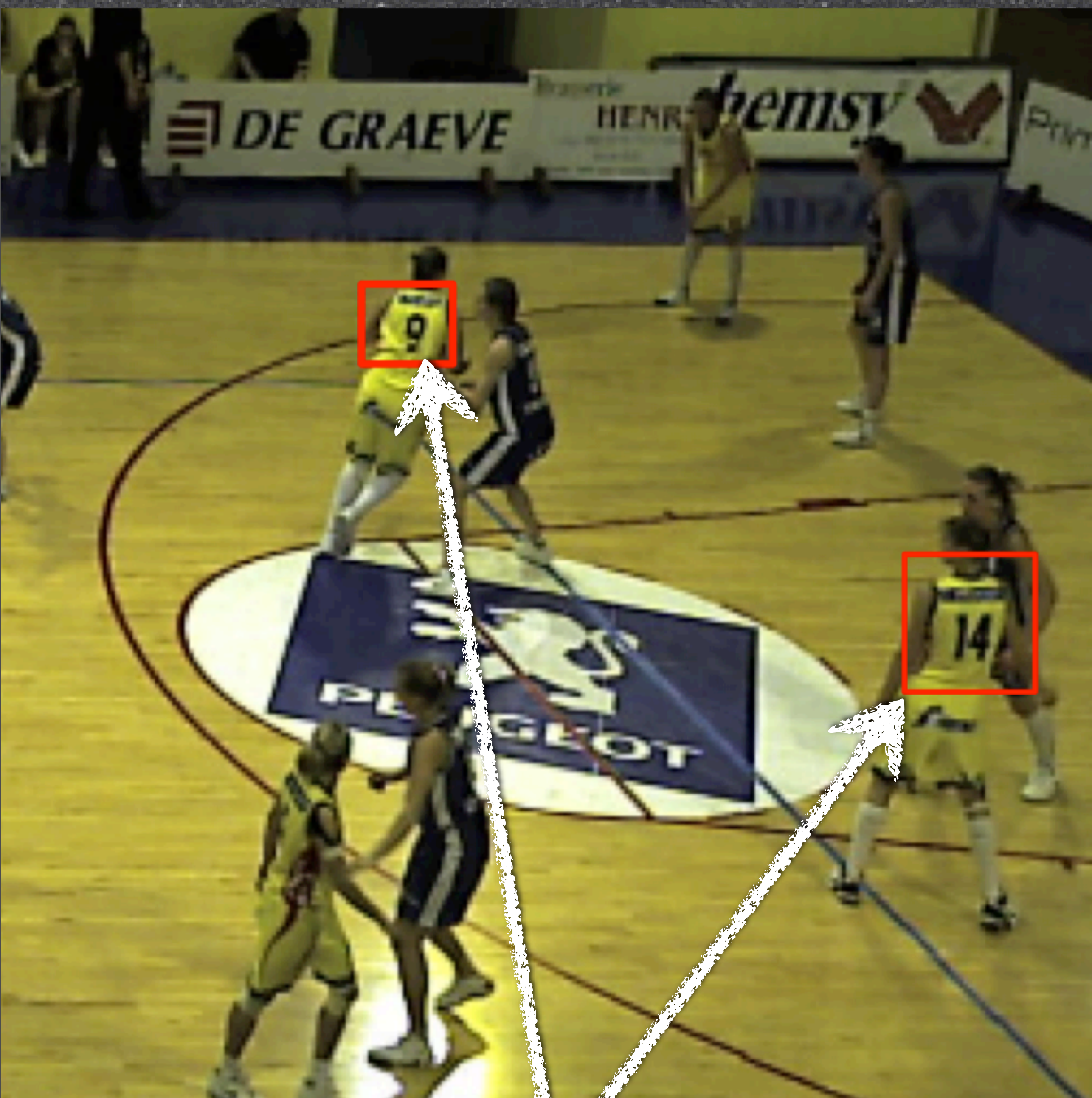
11

Specifically,



What if appearance features are not available every time?

Examples...



Color histogram is unreliable and hence is discarded.

Digit feature is available only when it faces the camera.

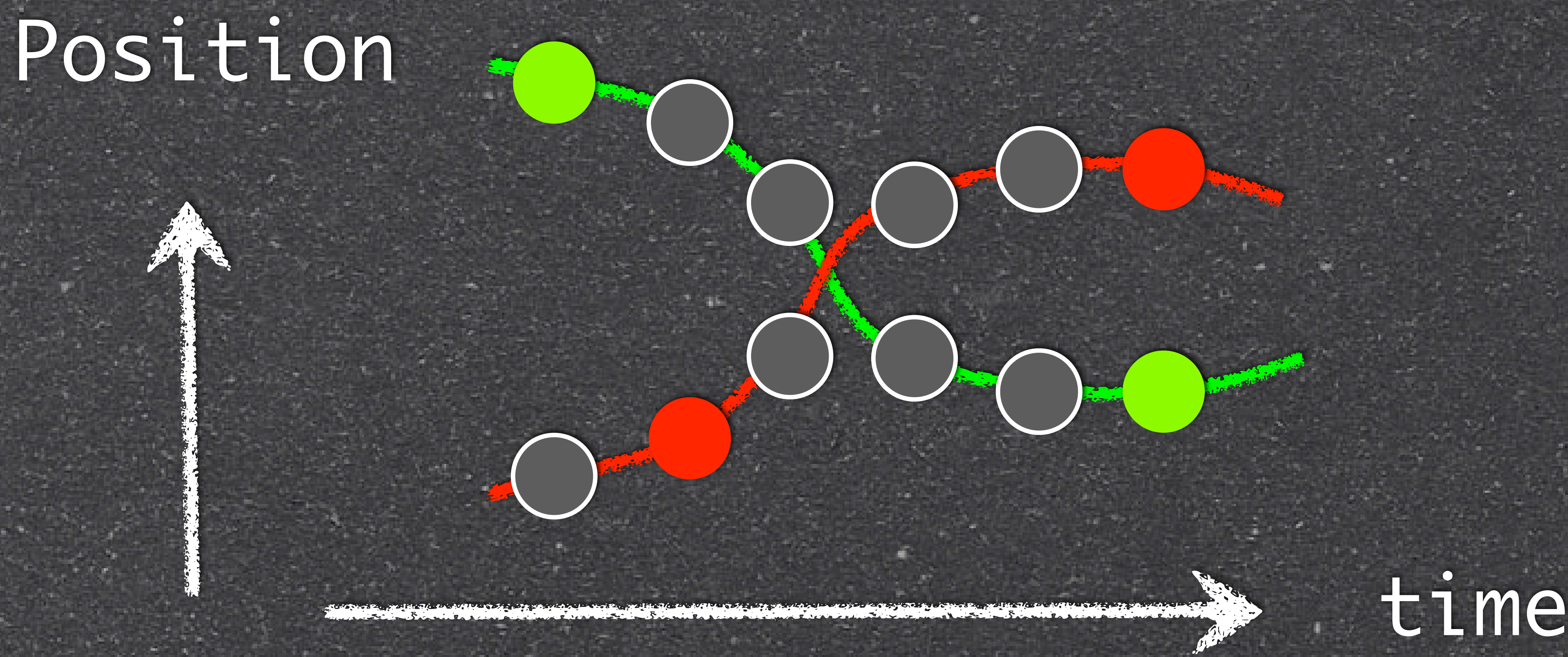
13

How to link the detections while exploiting such sporadic features?

- Construct a set of graphs
- Label the nodes consistently

14

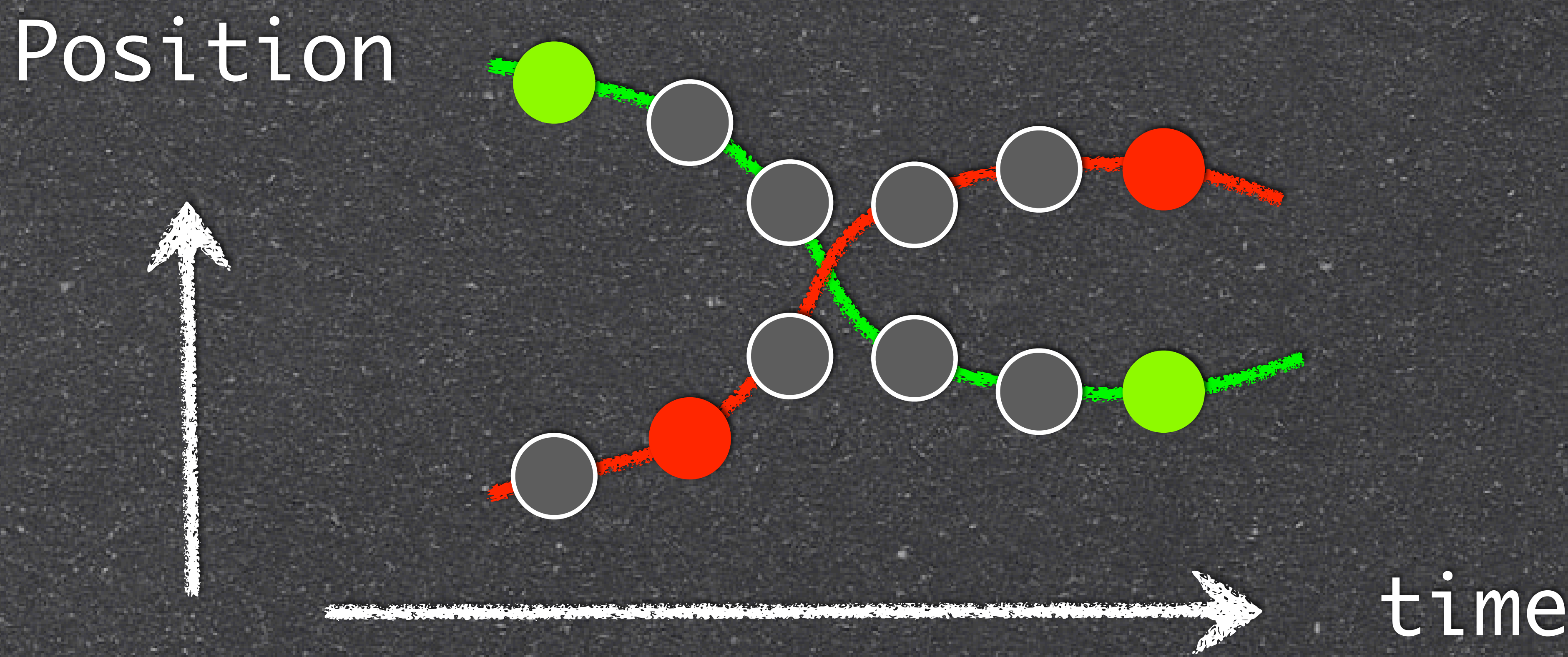
Construct a set of graphs...



How many different relationships between the nodes can you deduce?

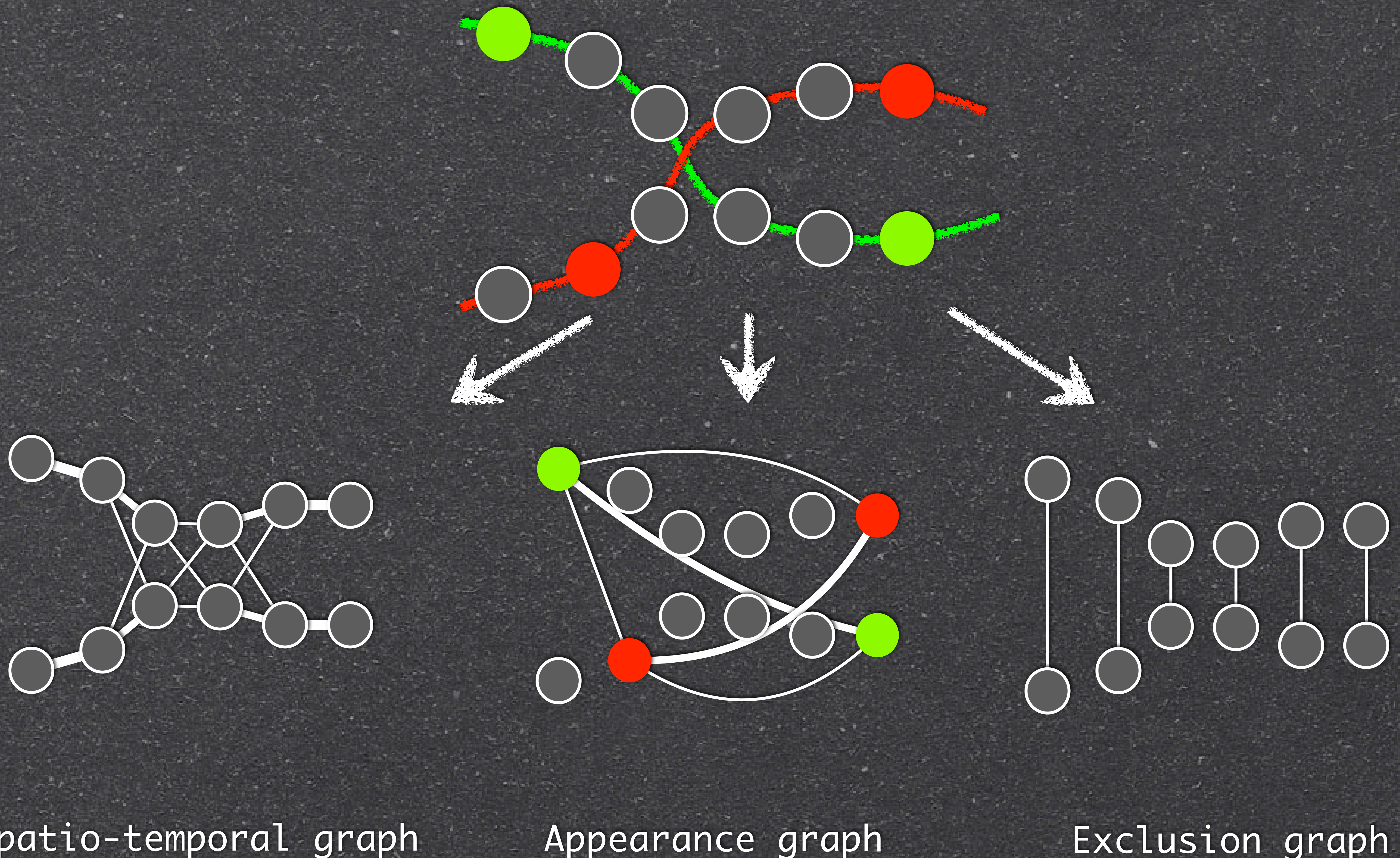
15

Construct a set of graphs...



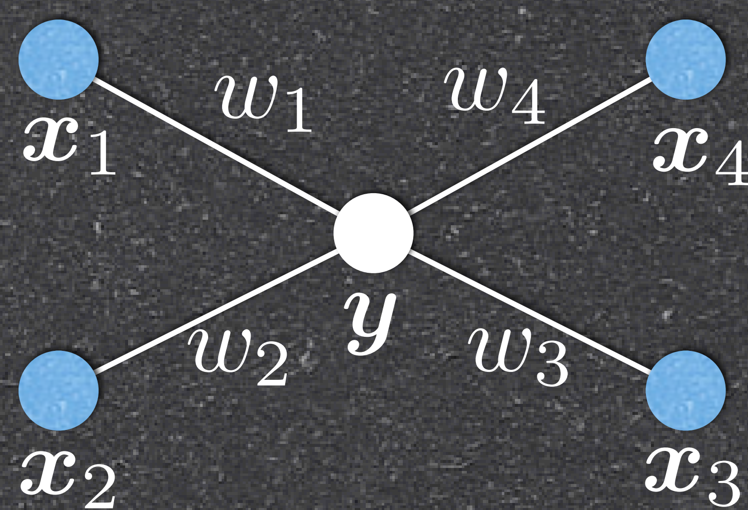
- Close/far in position-time
- Close/far in appearance(s)
- Two nodes that occur at the same time are different objects.

Construct a set of graphs...



Construct a set of graphs...

- Represent a sample by a linear combination of its local neighbors.
  - ~ Locally linear embedding (LLE)



$$\boldsymbol{y} \approx \sum_{i=1}^{4} w_i \boldsymbol{x}_i$$

- Solve for the “reconstruction” weights

$$\boldsymbol{w}^\star = \operatorname*{argmin}_{\boldsymbol{w}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 \text{ subject to } \mathbf{1}^\top \boldsymbol{w} = 1, \boldsymbol{w} \succeq \mathbf{0}.$$

Neighbors of $\boldsymbol{y}$ — Probability simplex

Construct a set of graphs...

- Represent a sample by a linear combination of its local neighbors.
  - ~ Locally linear embedding (LLE)

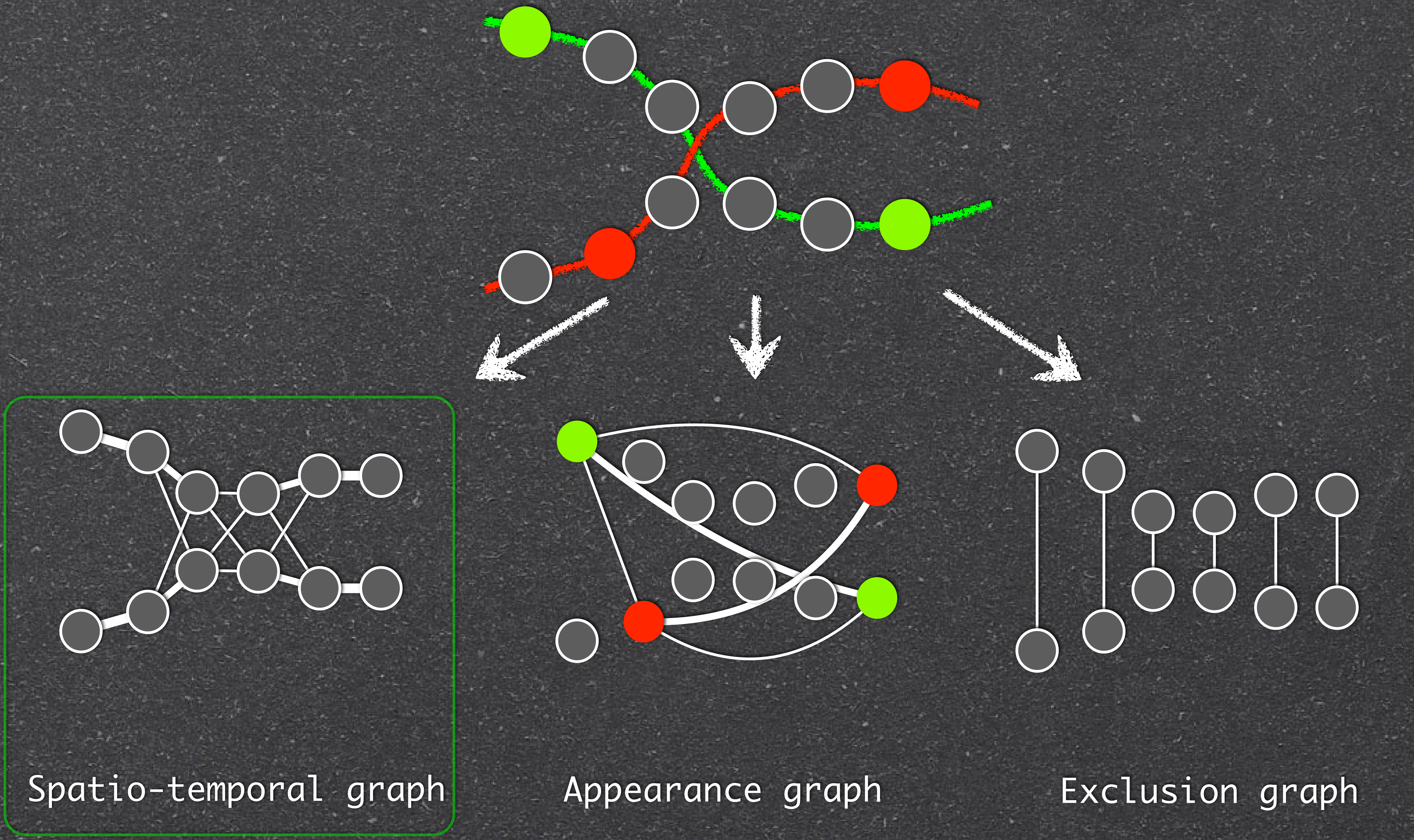
What is sample y?
How to define the neighbors X of y?

- Solve for the “reconstruction” weights

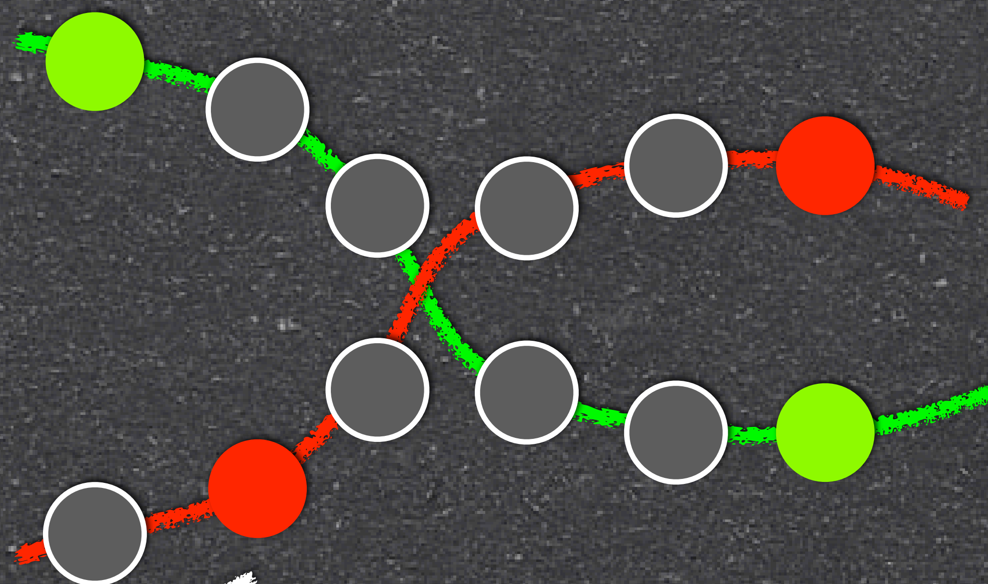
$$\boldsymbol{w}^\star = \operatorname*{argmin}_{\boldsymbol{w}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 \text{ subject to } \mathbf{1}^\top \boldsymbol{w} = 1, \boldsymbol{w} \succeq \mathbf{0}.$$

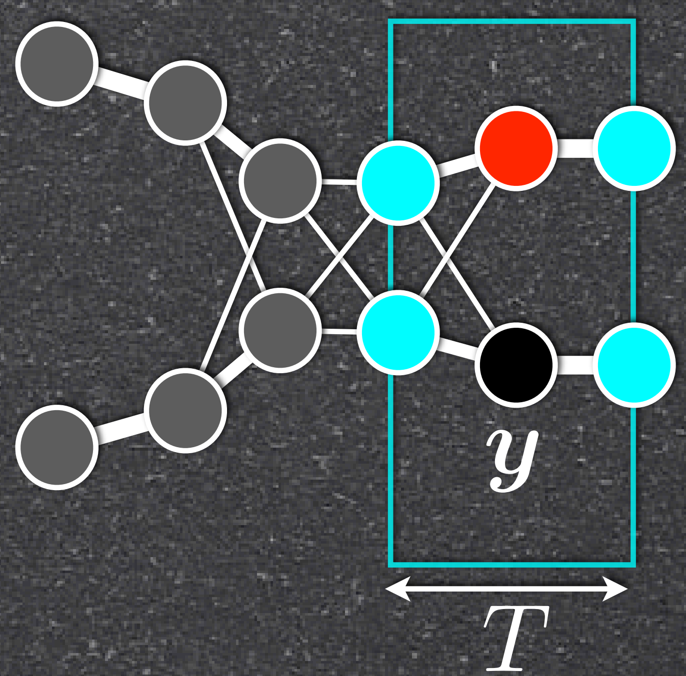
Neighbors of $\boldsymbol{y}$ — Probability simplex

Construct a set of graphs...



Construct a set of graphs...





Spatio-temporal graph

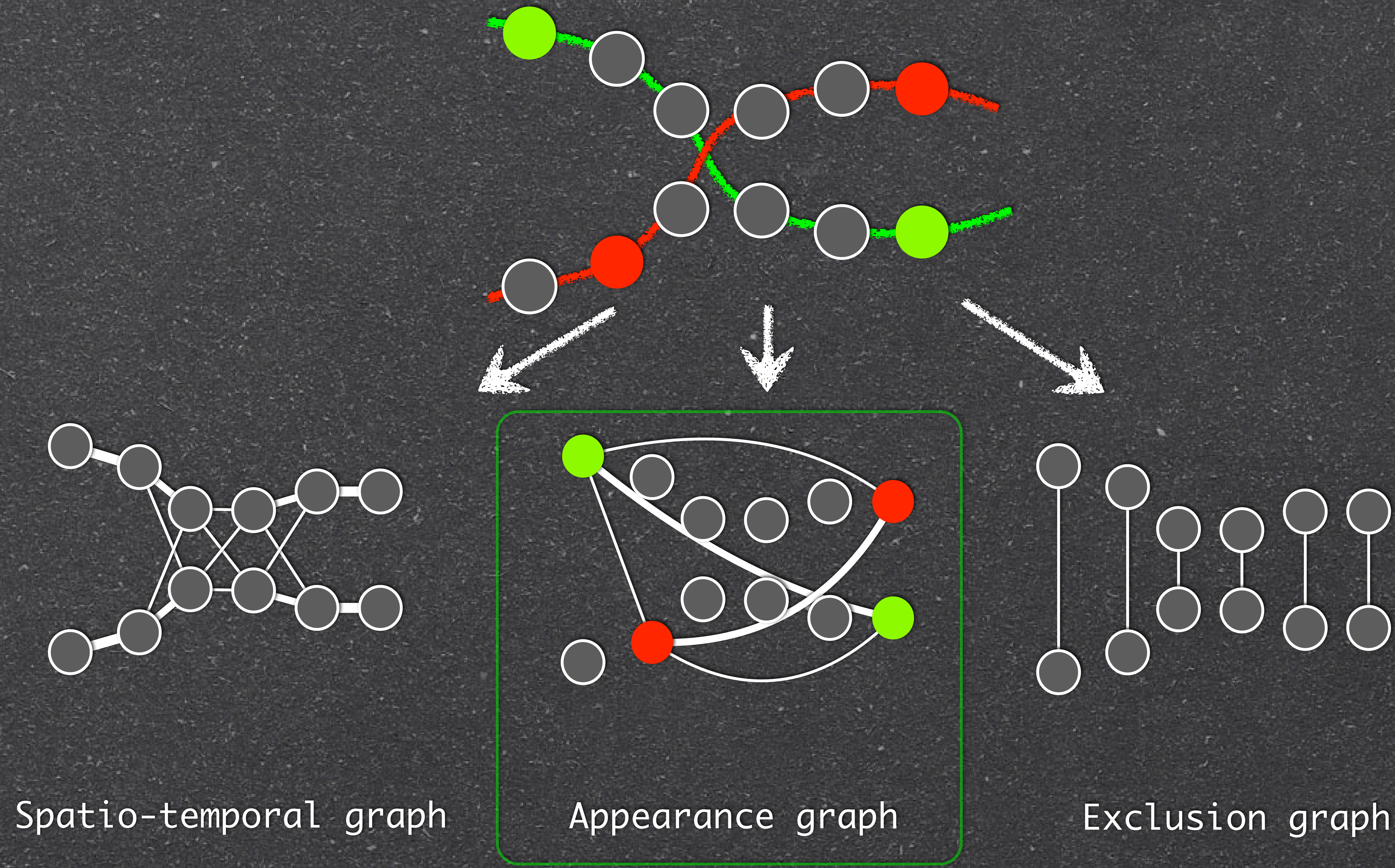
$$\boldsymbol{y} = (\gamma t, \boldsymbol{c})^\top$$

t: time instant
c: position (bounding box)

Relative importance between time difference and position difference

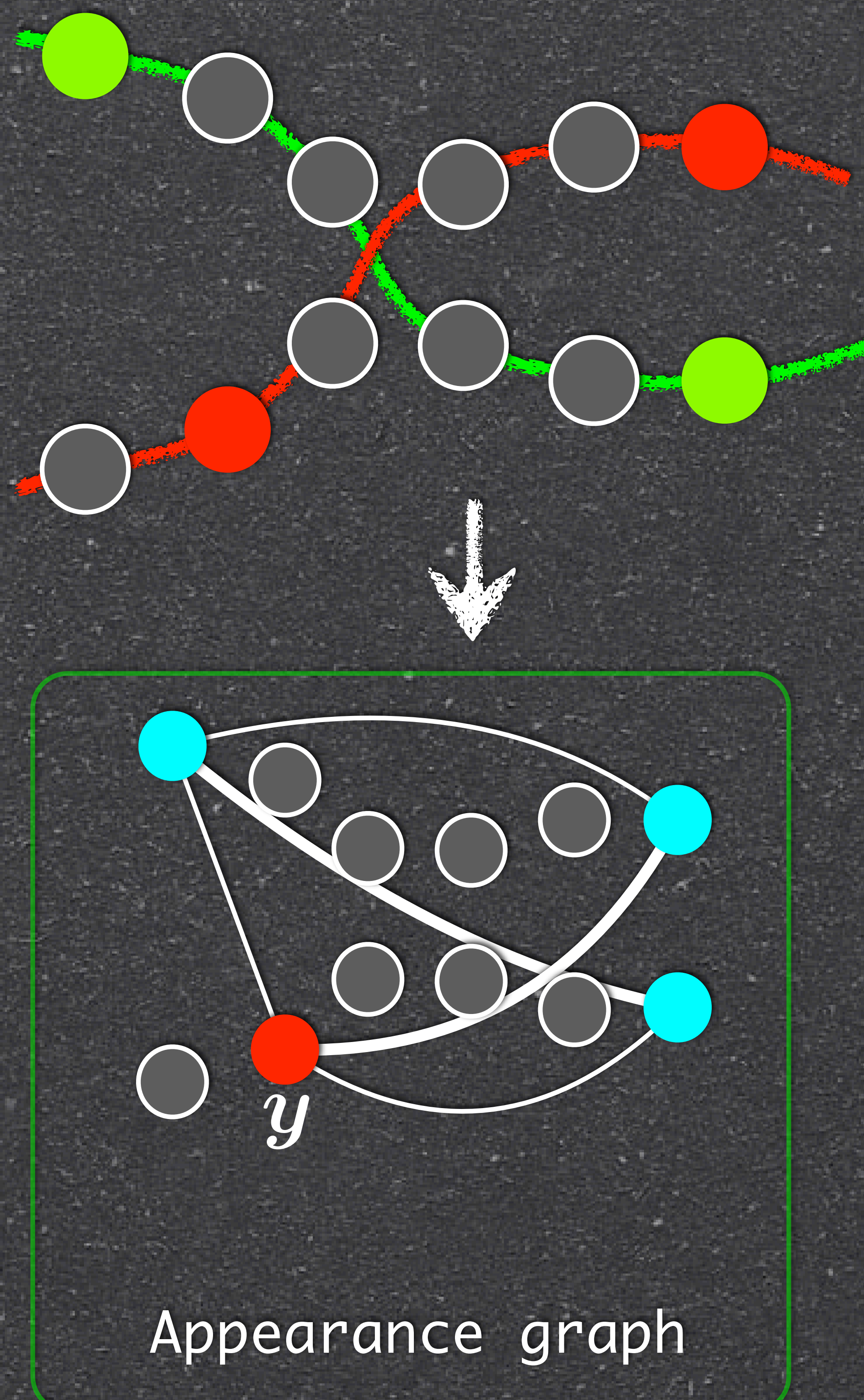
X: samples within T except those occurring at t

Construct a set of graphs...



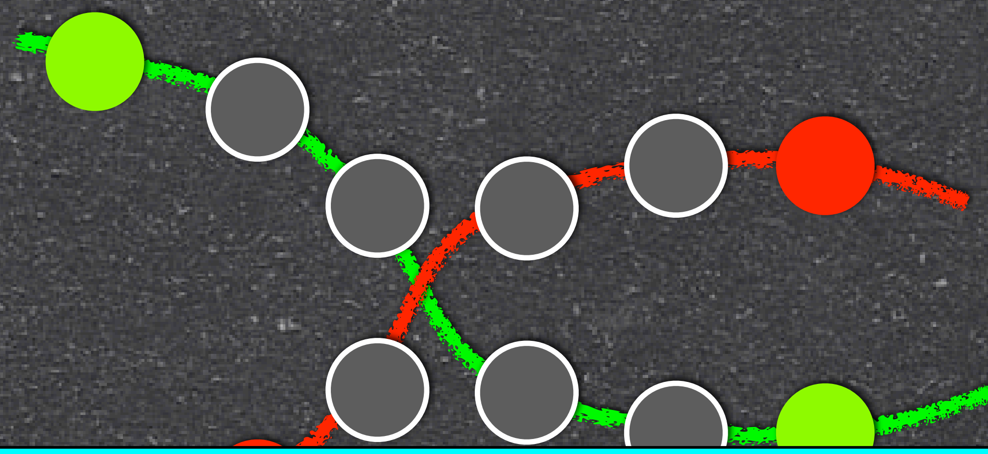
Construct a set of graphs...

y: appearance feature vector (e.g., color histogram)



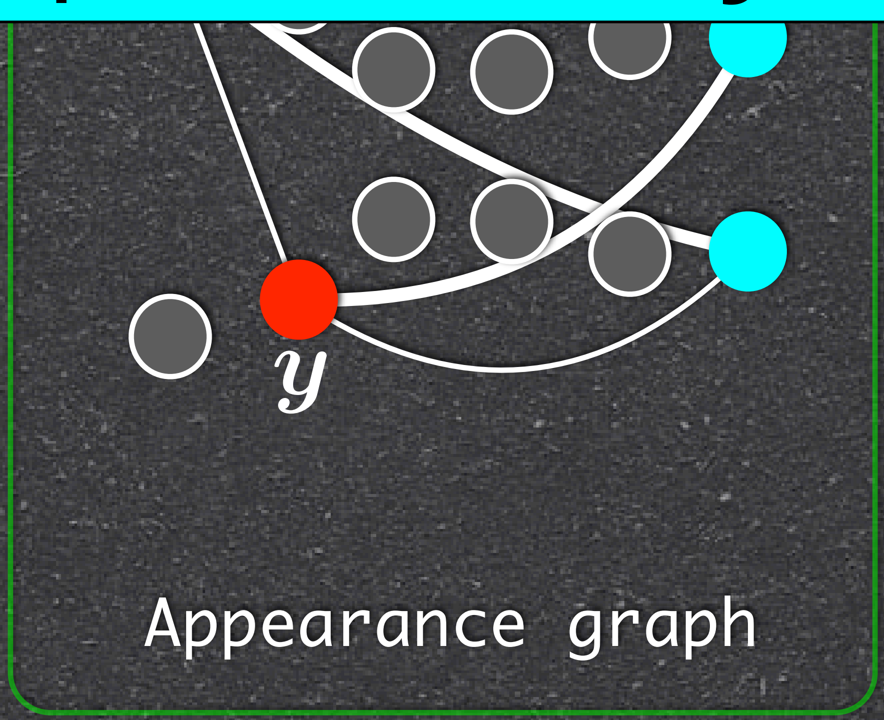
X: all other samples that have appearance features, except those occurring at t.

Construct a set of graphs...



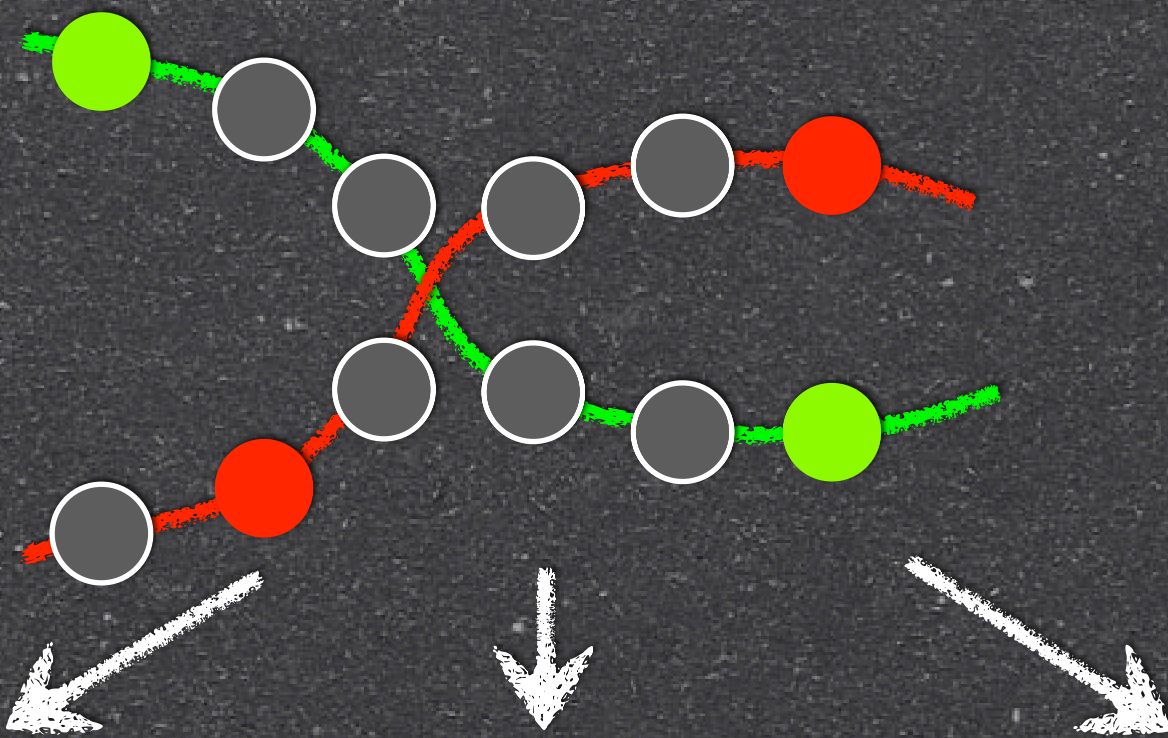
Allows connection between nodes for which appearance features occur only sporadically.

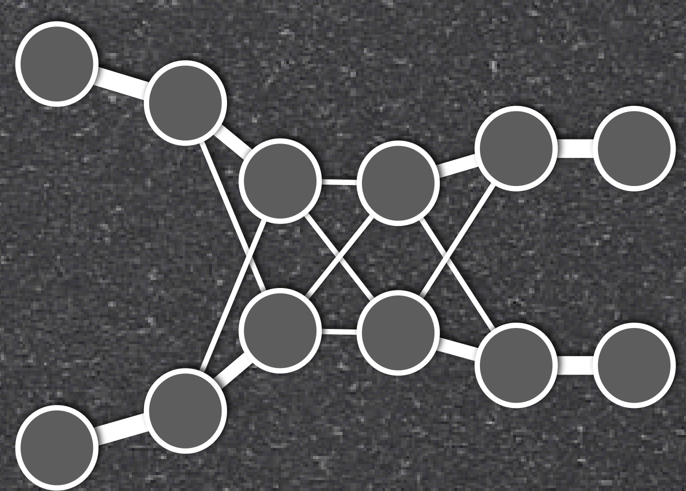
feature vector (e.g., color histogram)

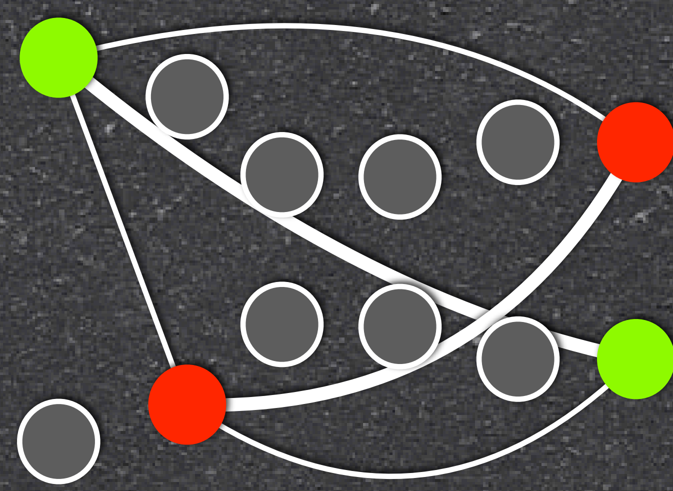


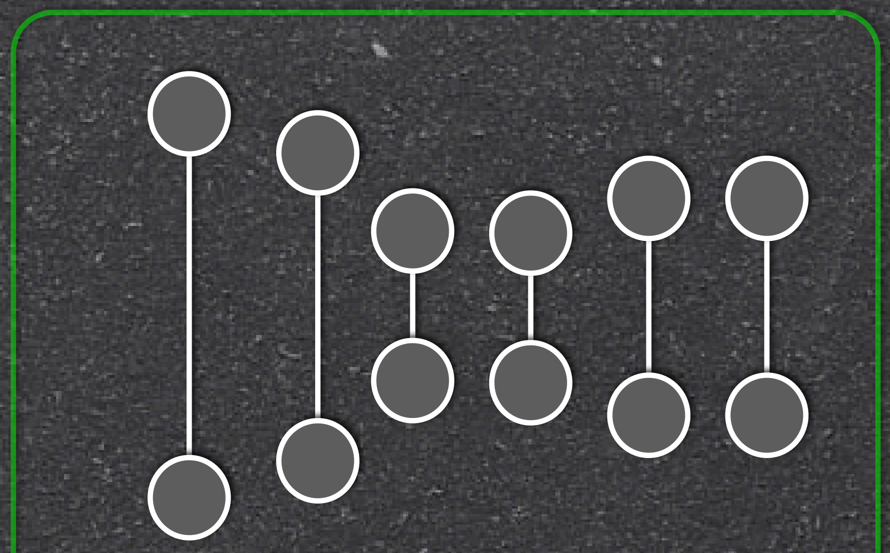
samples that have appearance features, except those occurring at t.

Construct a set of graphs...

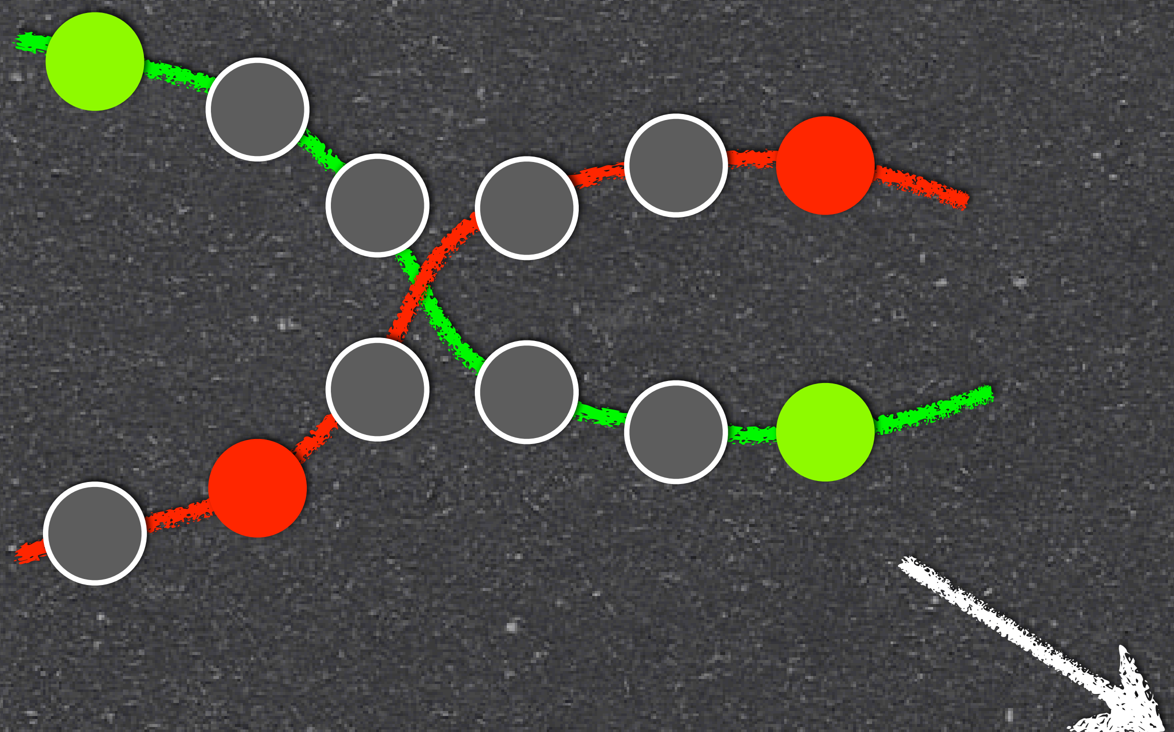




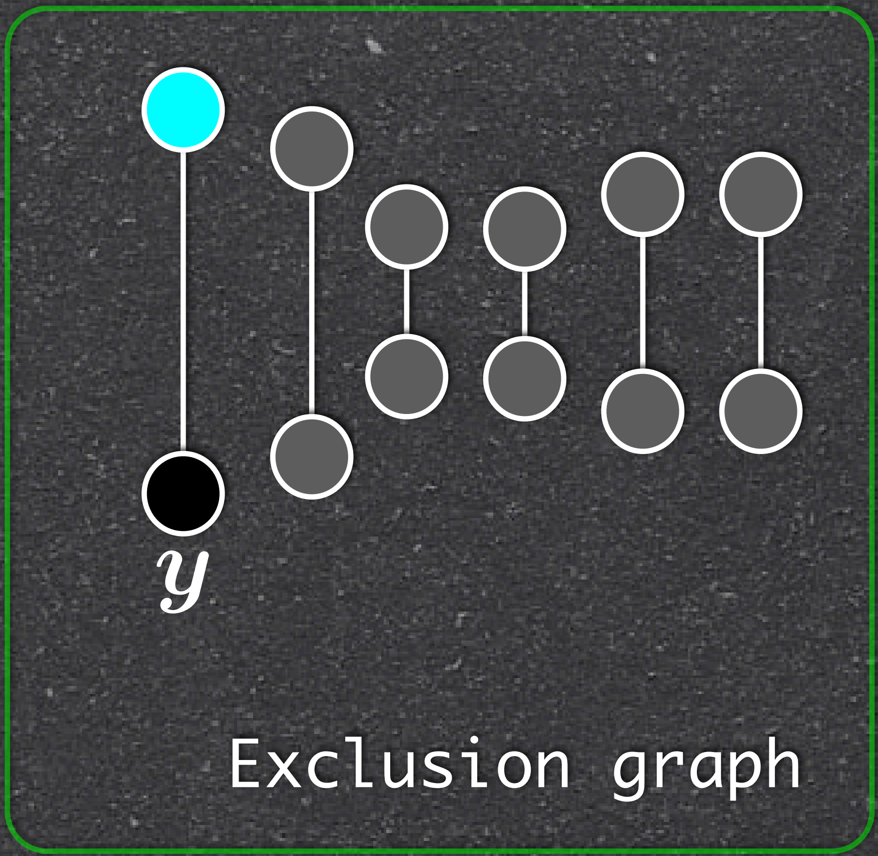




Construct a set of graphs...



- Exploit the <u>mutual exclusivity</u> between the nodes.
- X: all nodes that occur at the same time instant.



Recall..

- A set of graphs
- Spatio-temporal and appearance graph(s)
  - capture the proximity in space, time and appearance.
- Exclusion graph
  - co-occurring nodes CANNOT have same labels.

Multi-object tracking as a...

- Consistent labeling problem in a set of associated graphs.

28

Some preliminaries

$$G = (V, E, W)$$ Graph

$$\boldsymbol{z}_i$$ Label distribution of the i-th node

$$Z = (\boldsymbol{z}_1, ..., \boldsymbol{z}_{|V|})^\top$$ Label assignment matrix

$$\mathcal{E}_G(Z) = \frac{1}{2} \sum_{i=1}^{|V|} \sum_{j \in \mathcal{N}_i} W_{ij} \|\boldsymbol{z}_i - \boldsymbol{z}_j\|^2$$ Labeling error[1]

[1] X. Zhu, Z. Ghahramani, and J. Lafferty, “Semi-supervised learning using Gaussian fields and harmonic functions”, ICML, 2003.

Some preliminaries

$G = (V, E, W)$ Graph

$z_i(j)$	probability that the i-th node has a label 1 ≤ j ≤ |V|.	z_i	Label distribution of the i-th node

$Z = (z_1, ..., z_{|V|})^\top$ Label assignment matrix

$\mathcal{E}_G(Z) = \frac{1}{2} \sum_{i=1}^{|V|} \sum_{j \in \mathcal{N}_i} W_{ij} \|z_i - z_j\|^2$ Labeling error [1]

[1] X. Zhu, Z. Ghahramani, and J. Lafferty, “Semi-supervised learning using Gaussian fields and harmonic functions”, ICML, 2003.

30

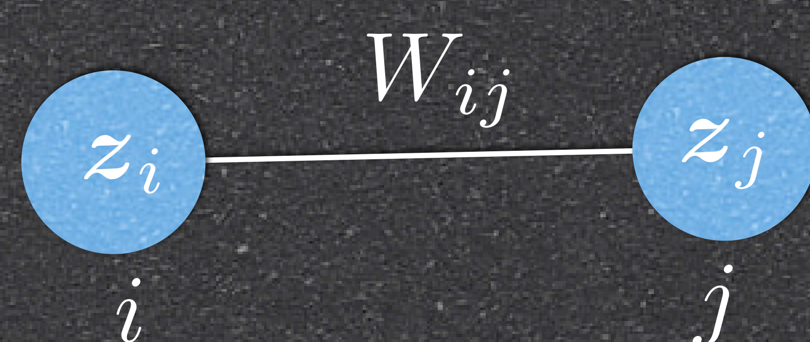
Some preliminaries

$$G = (V, E, W)$$ Graph

$$\boldsymbol{z}_i$$ Label distribution of the i-th node

$$Z = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{|V|})^\top$$ Label assignment matrix

$$\mathcal{E}_G(Z) = \frac{1}{2} \sum_{i=1}^{|V|} \sum_{j \in \mathcal{N}_i} W_{ij} \|\boldsymbol{z}_i - \boldsymbol{z}_j\|^2$$ Labeling error [1]



[1] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions", ICML, 2003.

31

Some preliminaries

$G = (V, E, W)$ Graph

$\boldsymbol{z}_i$ Label distribution of the i-th node

$Z = (\boldsymbol{z}_1, ..., \boldsymbol{z}_{|V|})^\top$ Label assignment matrix

$$\mathcal{E}_G(Z) = \frac{1}{2} \sum_{i=1}^{|V|} \sum_{j \in \mathcal{N}_i} W_{ij} \|\boldsymbol{z}_i - \boldsymbol{z}_j\|^2$$

$$= \mathbf{Tr}(Z^\top L Z)$$

Labeling error[1]

[1] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions", ICML, 2003.

32

Some preliminaries

Graph Laplacian

- Positive semi-definite
- Labeling error is convex.

$$L = D - W$$

$$D = \mathbf{diag}(d_1, ..., d_{|V|})$$

$$d_i = \sum_{j \in \mathcal{N}_i} W_{ij}$$

$$= 1$$

raph

abel distribution of
he i-th node

abel assignment matrix

abeling error [1]

$$\mathcal{E}_G(Z) = \frac{1}{2} \sum_{i=1}^{|V} \sum_{j \in \mathcal{N}_i}$$

$$= \mathbf{Tr}(Z^\top L Z)$$

[1] X. Zhu, Z. Ghahramani, and J. Lafferty, “Semi-supervised learning using Gaussian fields and harmonic functions”, ICML, 2003.

So far,

- Laplacian L is used to denote a graph.
- Labeling error measures the inconsistency in labels between the neighboring nodes
- Labeling error in CONVEX.

In our framework,

Laplacian(s)

- 1 spatio-temporal graph $L_0^{(+)}$
- K appearance graph(s) $L_p^{(+)}, p = 1, ..., K$
- 1 exclusion graph $L^{(-)}$

- We would like to:
  - minimize error due to spatio-temporal and appearance graphs
  - maximize error due to exclusion graph

$$Z^\star = \underset{Z \in \mathcal{P}}{\operatorname{argmin}} \sum_{p=0}^{K} \alpha_p \mathbf{Tr}(Z^\top L_p^{(+)} Z) - \mathbf{Tr}(Z^\top L^{(-)} Z)$$

weight to p-th graph

$$Z^\star = \operatorname*{argmin}_{Z \in \mathcal{P}} \sum_{p=0}^{K} \alpha_p \mathbf{Tr}(Z^\top L_p^{(+)} Z) - \mathbf{Tr}(Z^\top L^{(-)} Z)$$

Set of all row-stochastic matrices

Labeling error due to the p-th graph

Labeling error due to the exclusion graph weight to p-th graph

$$Z^\star = \operatorname*{argmin}_{Z \in \mathcal{P}} \sum_{p=0}^{K} \alpha_p \mathbf{Tr}(Z^\top L_p^{(+)} Z) - \mathbf{Tr}(Z^\top L^{(-)} Z)$$

Set of all row-stochastic matrices

Labeling error due to the p-th graph

Labeling error due to the exclusion graph

$$:= \operatorname*{argmin}_{Z \in \mathcal{P}} [f(Z) - g(Z)]$$

Difference of convex functions![2]

[2] B. K. Sriperumbudur and G. R. G. Lankriet, “On the convergence of the concave-convex procedure”, NIPS, 2009

Iterative solution

Randomly initialize $Z^{(0)} \in \mathcal{P}$

Set $k = 0$

Repeat

- Linearize $g(Z)$ at current solution $Z^{(k)}$
- Solve the convex optimization problem
  $$Z^{(k+1)} = \operatorname*{argmin}_{Z \in \mathcal{P}} \left[f(Z) - \nabla g^\top (Z^{(k)}) Z \right]$$
- Set $k = k + 1$

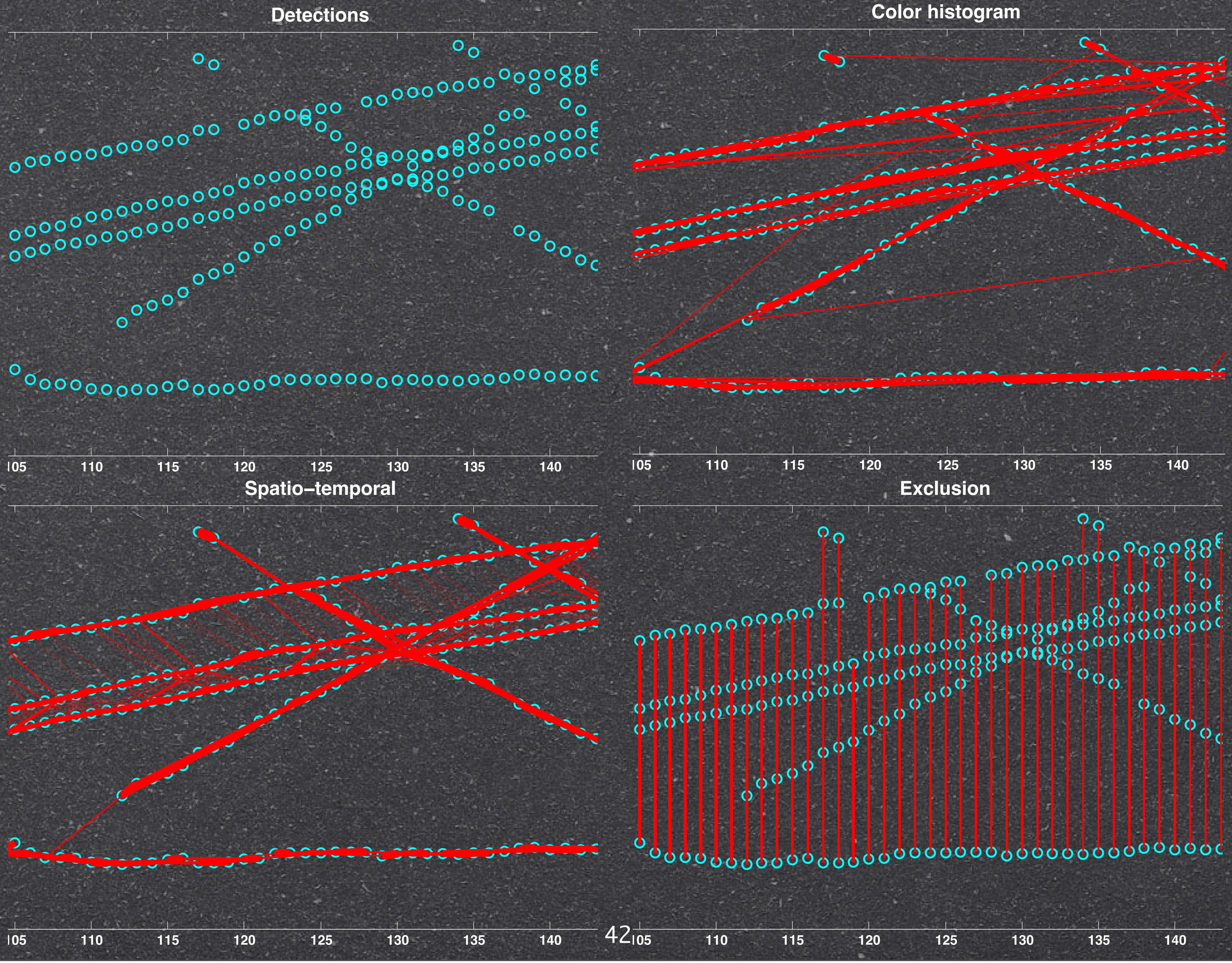
Until $\|Z^{(k+1)} - Z^{(k)}\|_F < \epsilon$

Some results

- Three datasets
  - APIDIS (multi-view, basketball)
  - PETS2009 S2/L1 (monocular, surveillance)
  - TUD Stadtmitte (monocular, surveillance)

40

Results


Detections
Color histogram
Spatio−temporal
Exclusion
105
110
115
120
125
130
135
140

Some numerical results

- Multiple Object Tracking Accuracy (MOTA)
- Multiple Object Tracking Precision (MOTP)

43

Some numerical results

Measures # errors in tracking

↓

- Multiple Object Tracking Accuracy (MOTA)
- Multiple Object Tracking Precision (MOTP)

↑

Measures how well the detection is aligned with the ground truth

Some numerical results

DATASET	No Appearance		With Appearance	
	MOTA	MOTP	MOTA	MOTP
TUD Stadtmitte	62.6	73.5	79.3	73.9
PETS 2009 S2/L1	82.75	71.21	91.01	70.99
APIDIS	81.25	57.13	83.80	60.01

Thank you!

Any questions?