Detection-based Multi-Object Tracking with Unreliable Appearance Features

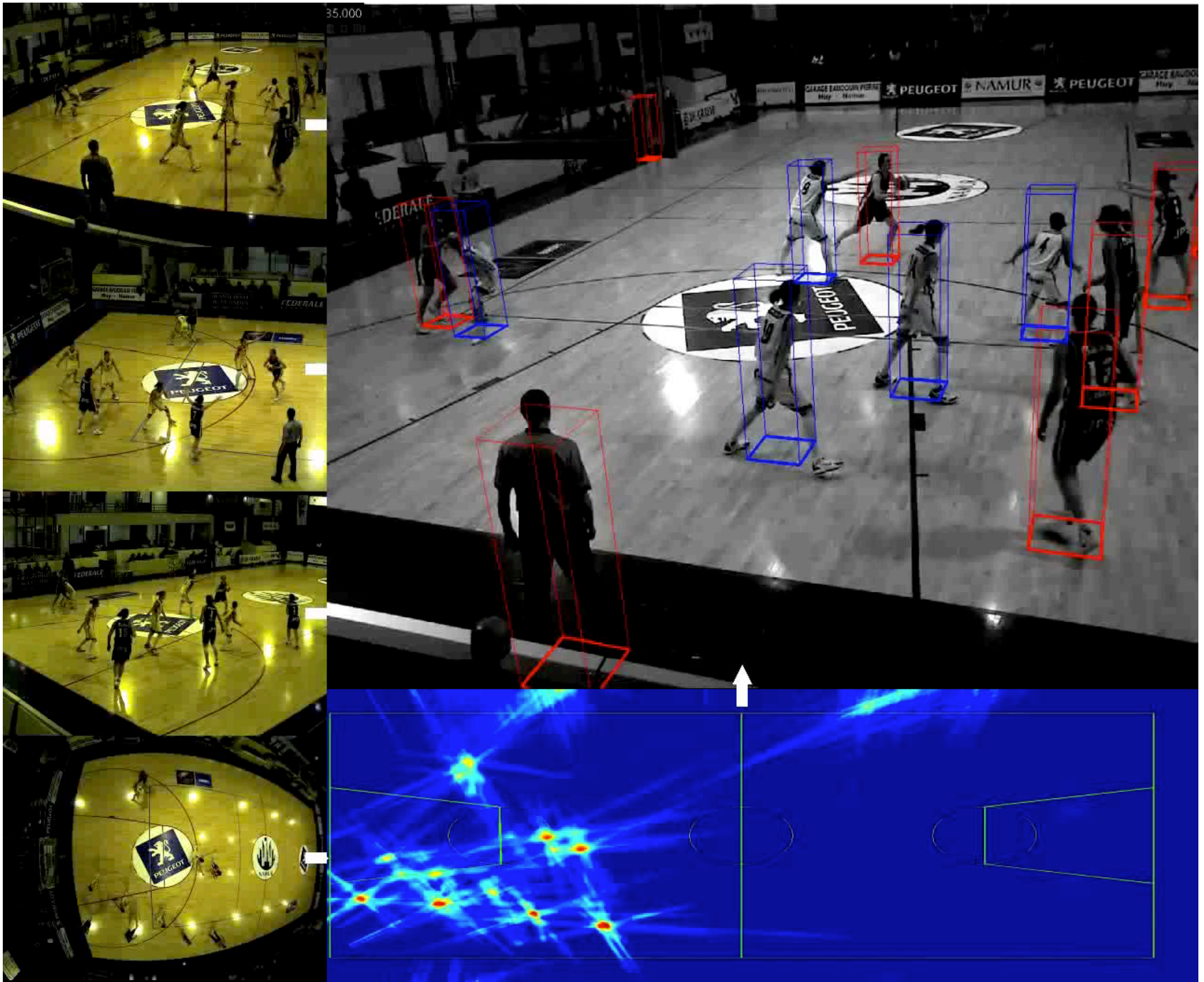
Public Defense

Amit Kumar K.C.

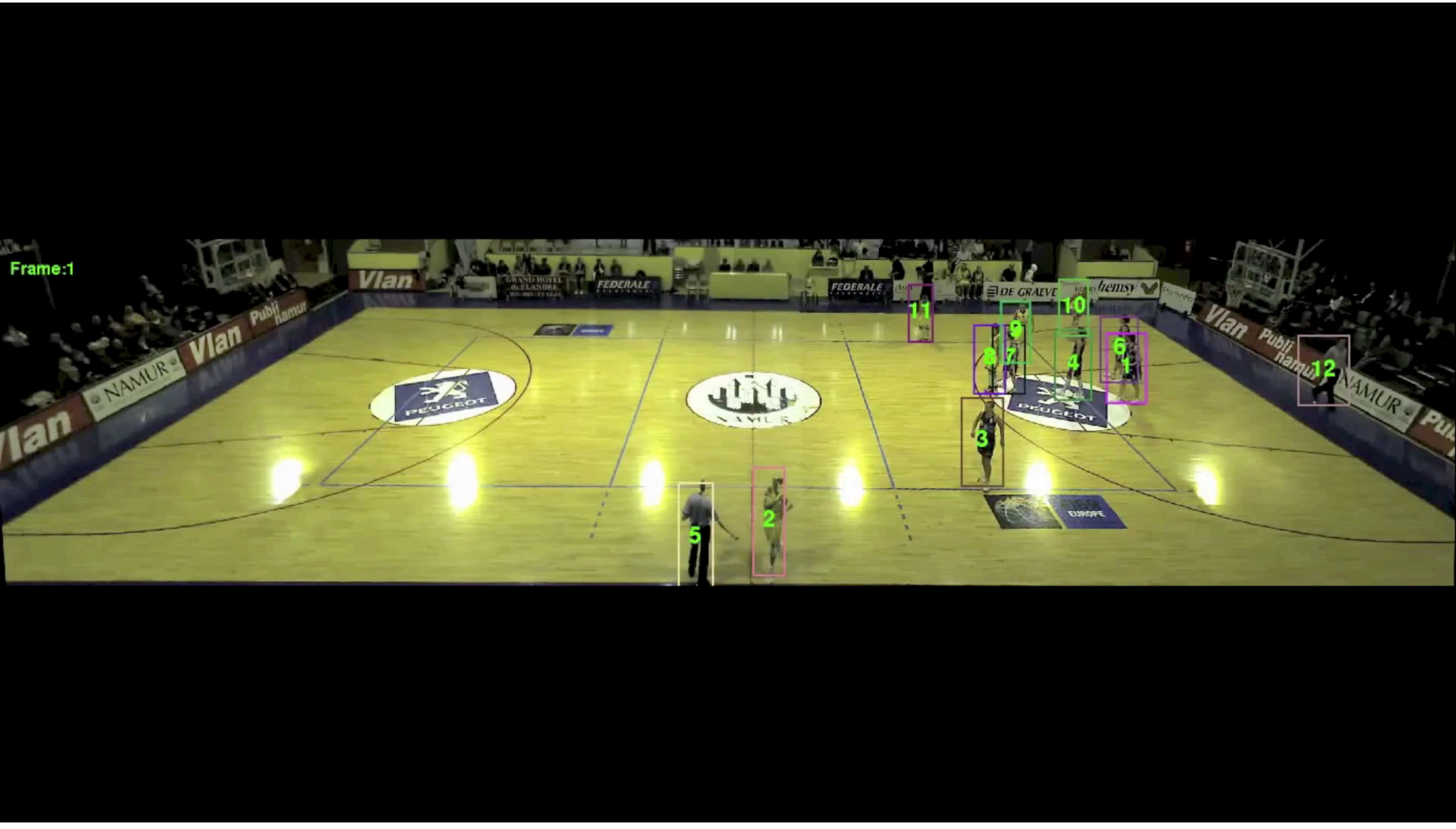
Thesis Committee

Christophe De Vleeschouwer (UCL) *Advisor*
Laurent Jacques (UCL) *Secretary*
Jean-François Delaigle (COFELY, Belgium)
Andrea Cavallaro (QMUL, UK)
François Fleuret (IDIAP/EPFL, Switzerland)
David Bol (UCL) *President*

Detections



Tracking results

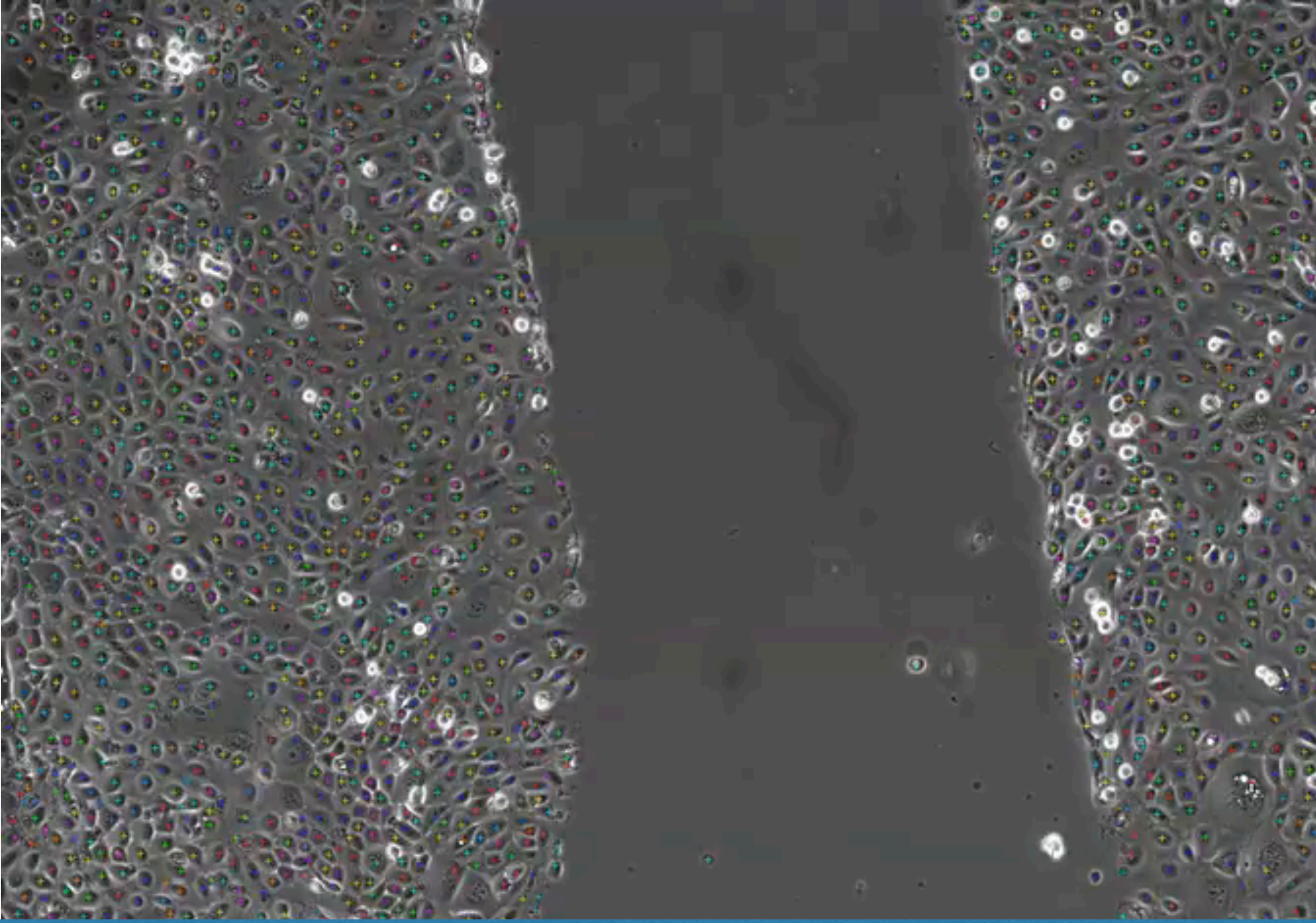


Recognition results

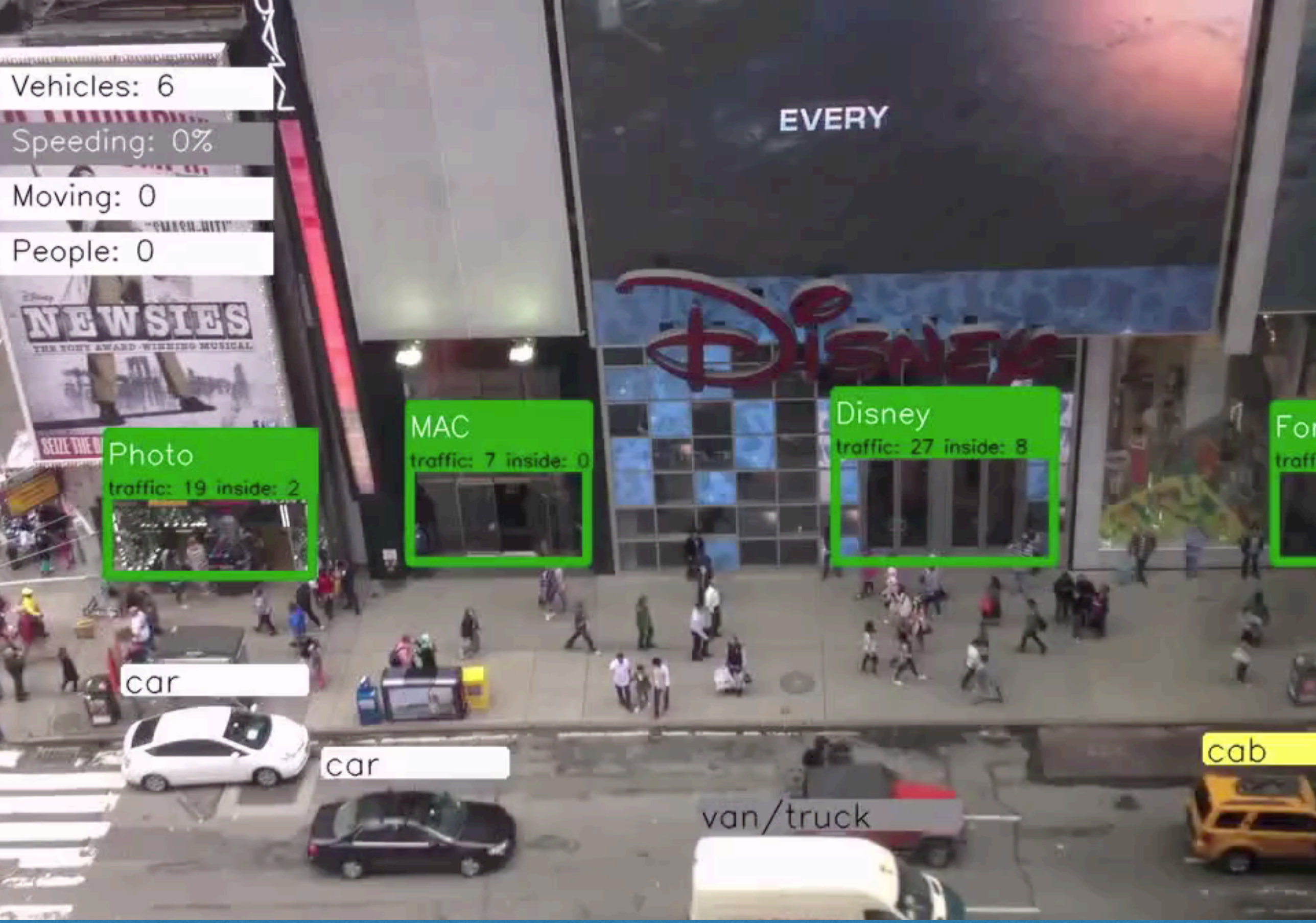
Organization

- Detection-based MOT
- Limitation of previous arts
- Contributions
- Validations
- Conclusion and future directions

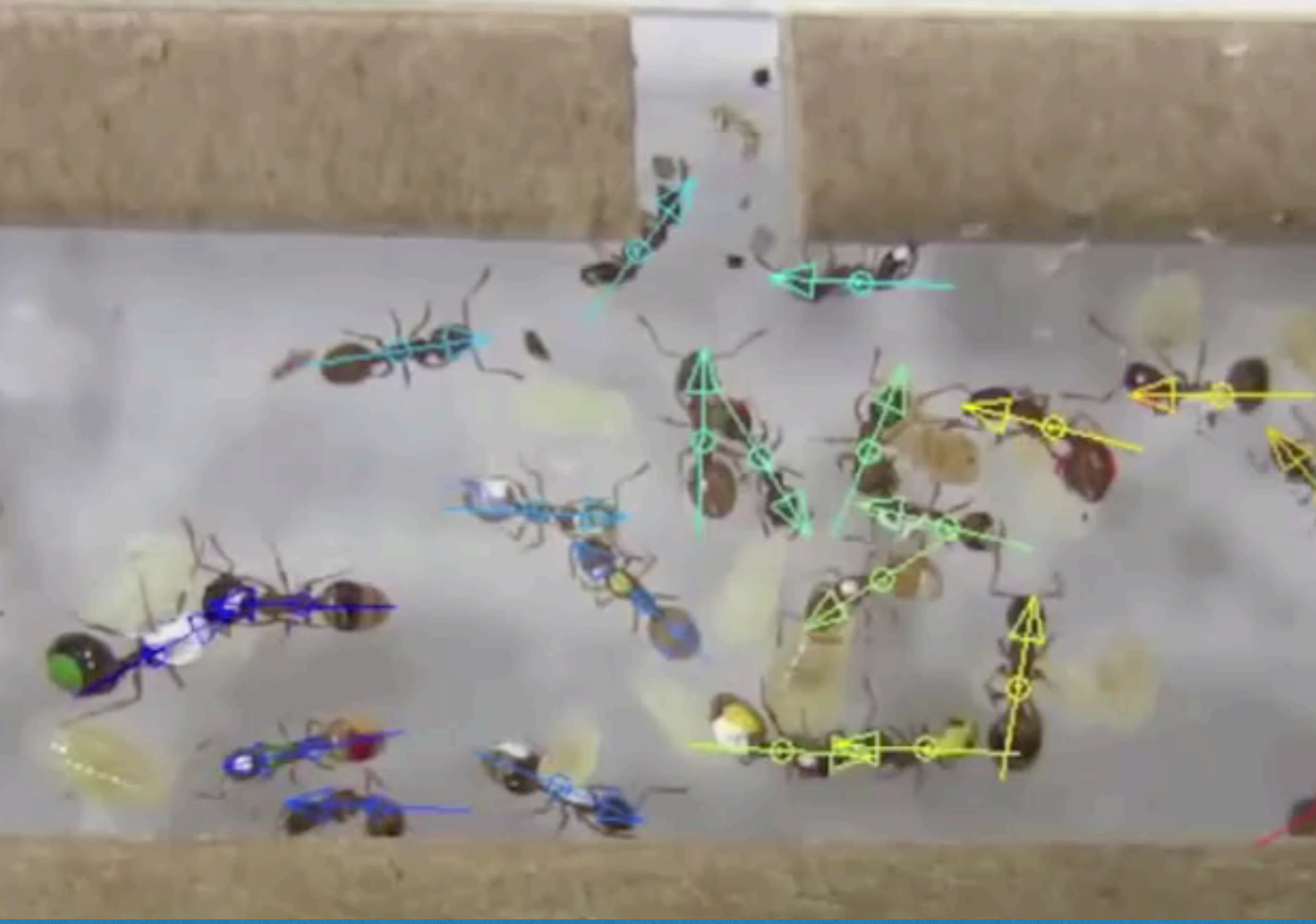
Multi-object tracking (MOT) is **ubiquitous**.



Cell migration during wound healing



Surveillance

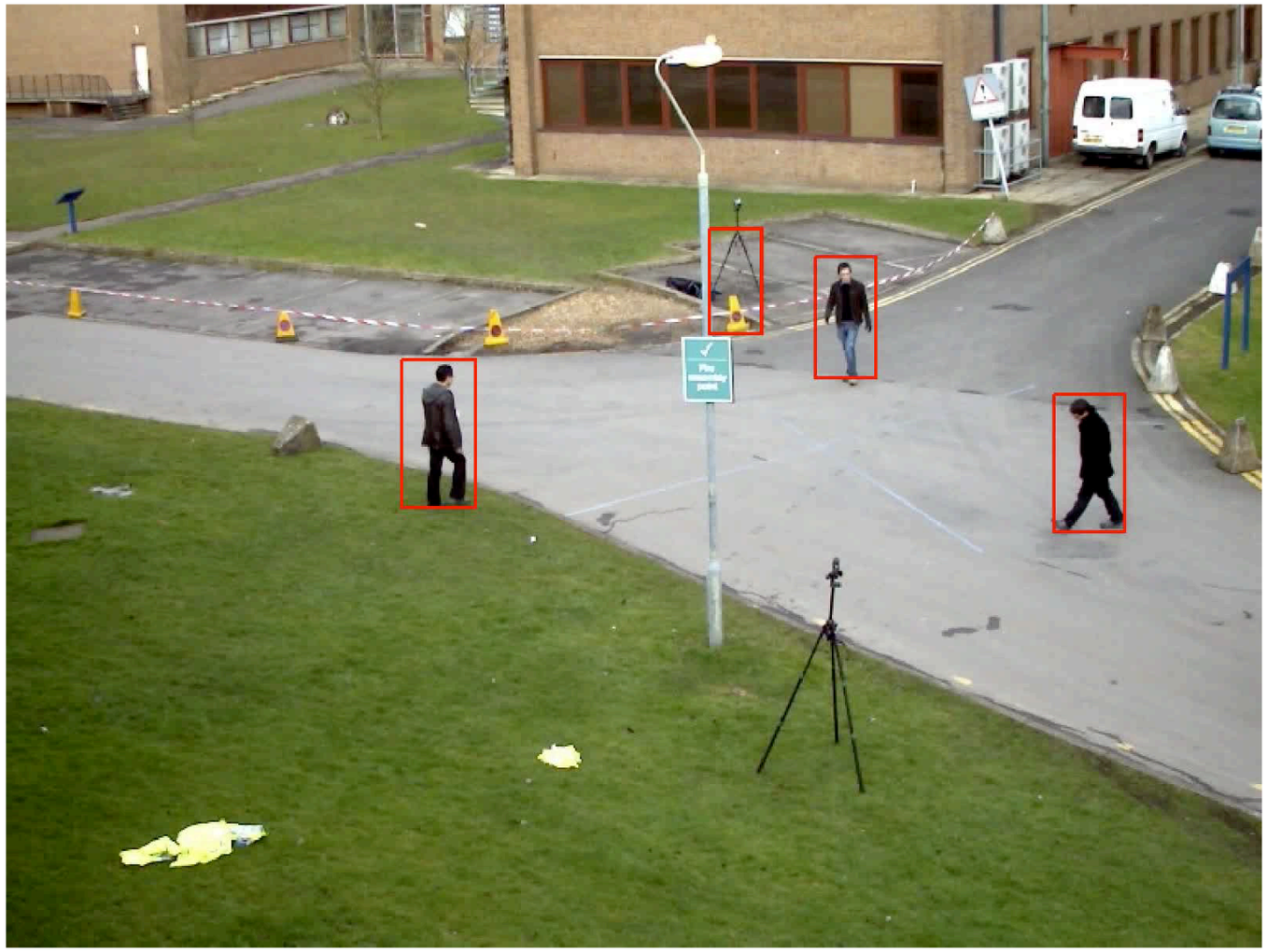


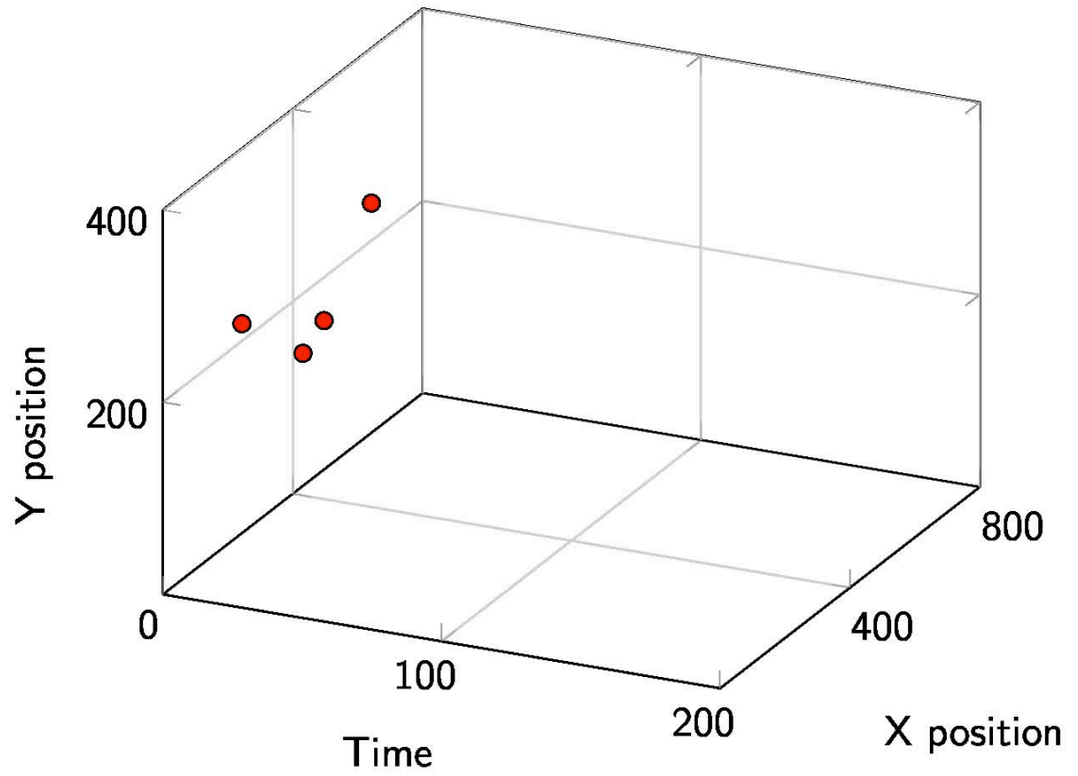
Insect behavior

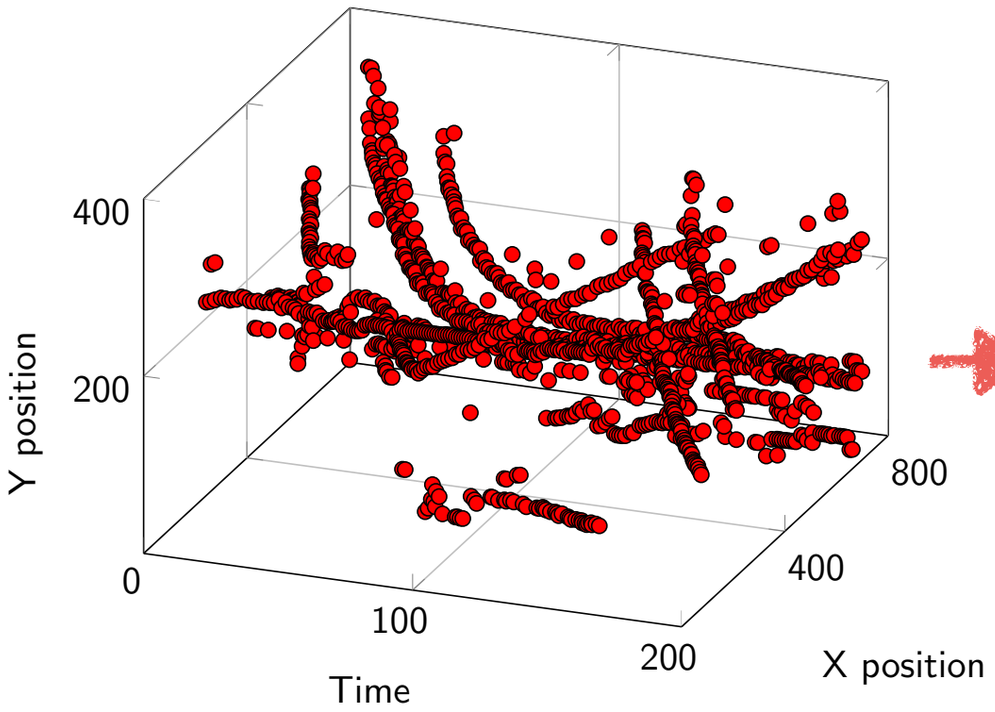
Multi-object tracking is ubiquitous.

Sport analysis

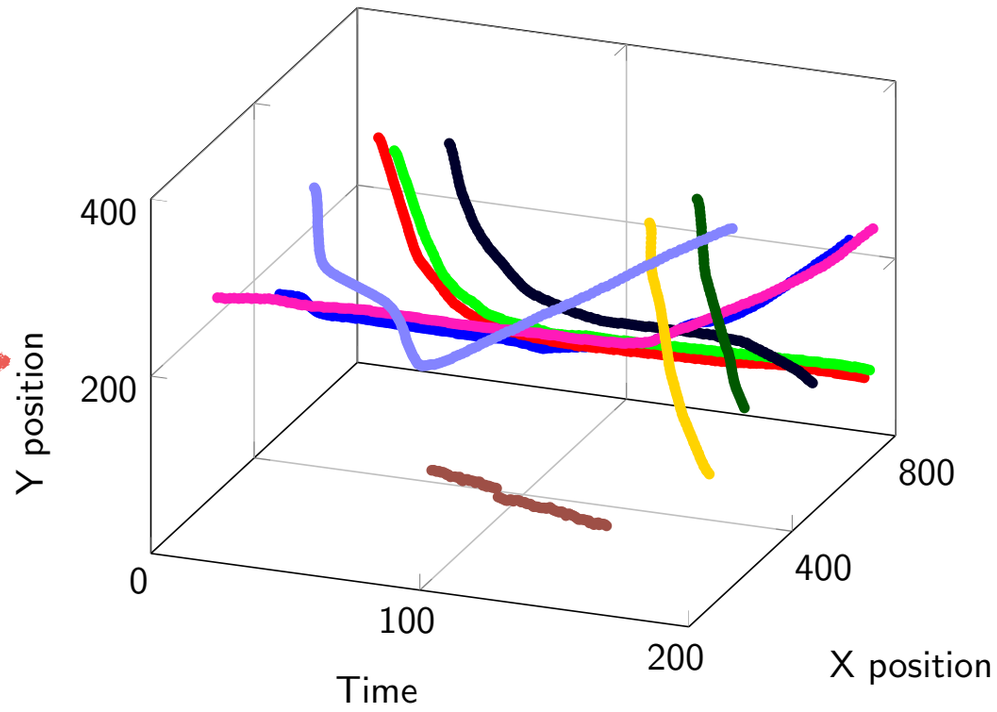
Detection-based MOT




400
200
Y position
0
100
200
Time
400
800
X position

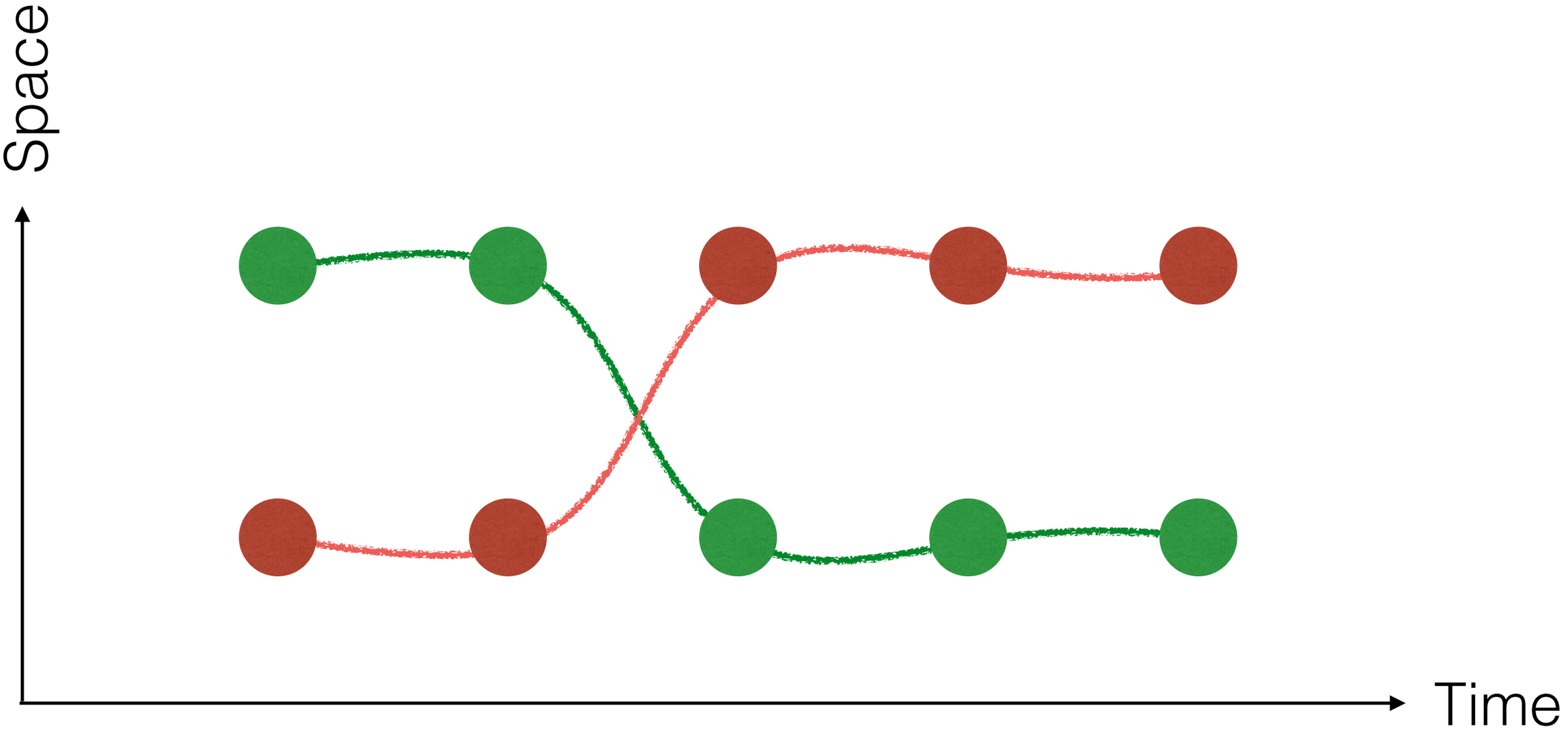


Data association

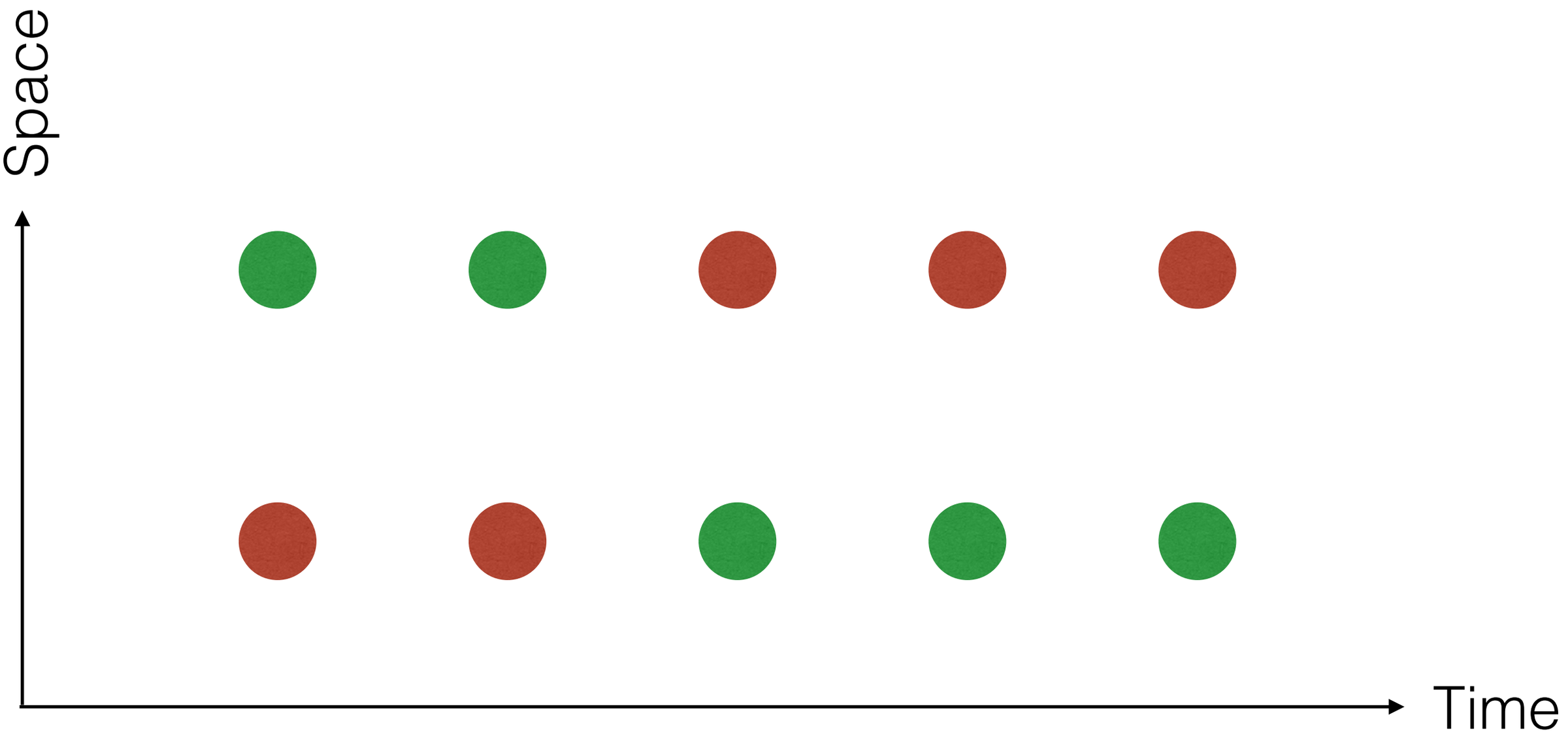


Previous arts in detection- based MOT

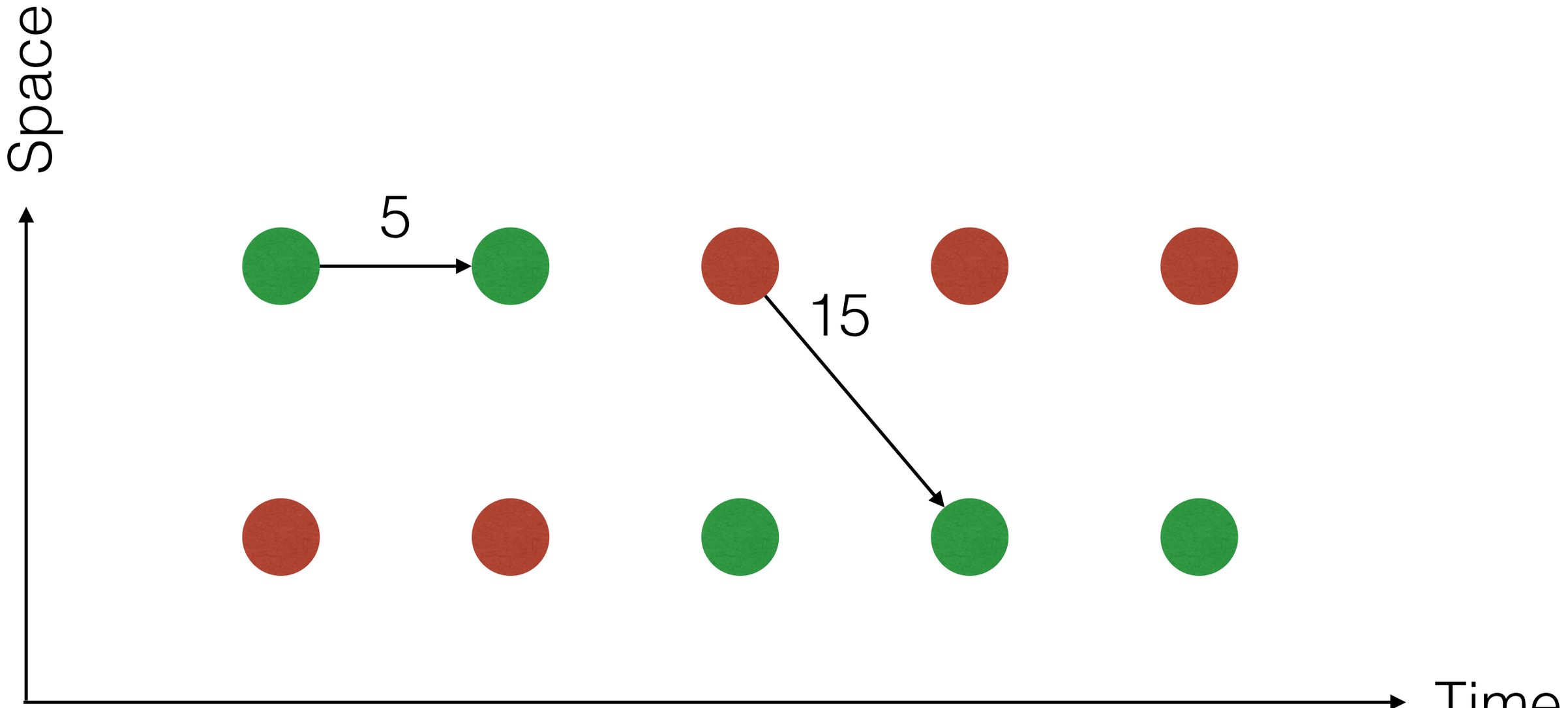
Two targets



Detections

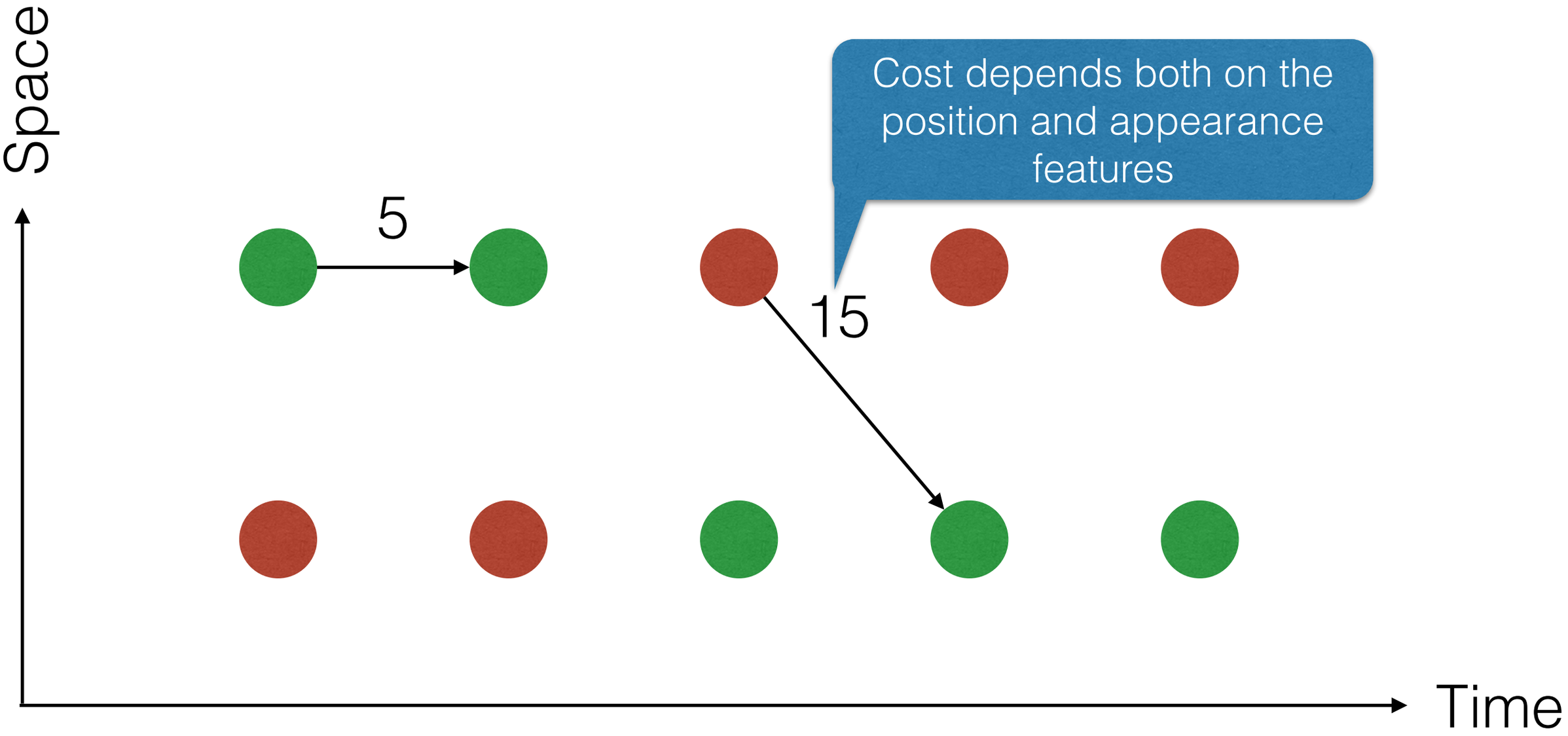


Graph construction



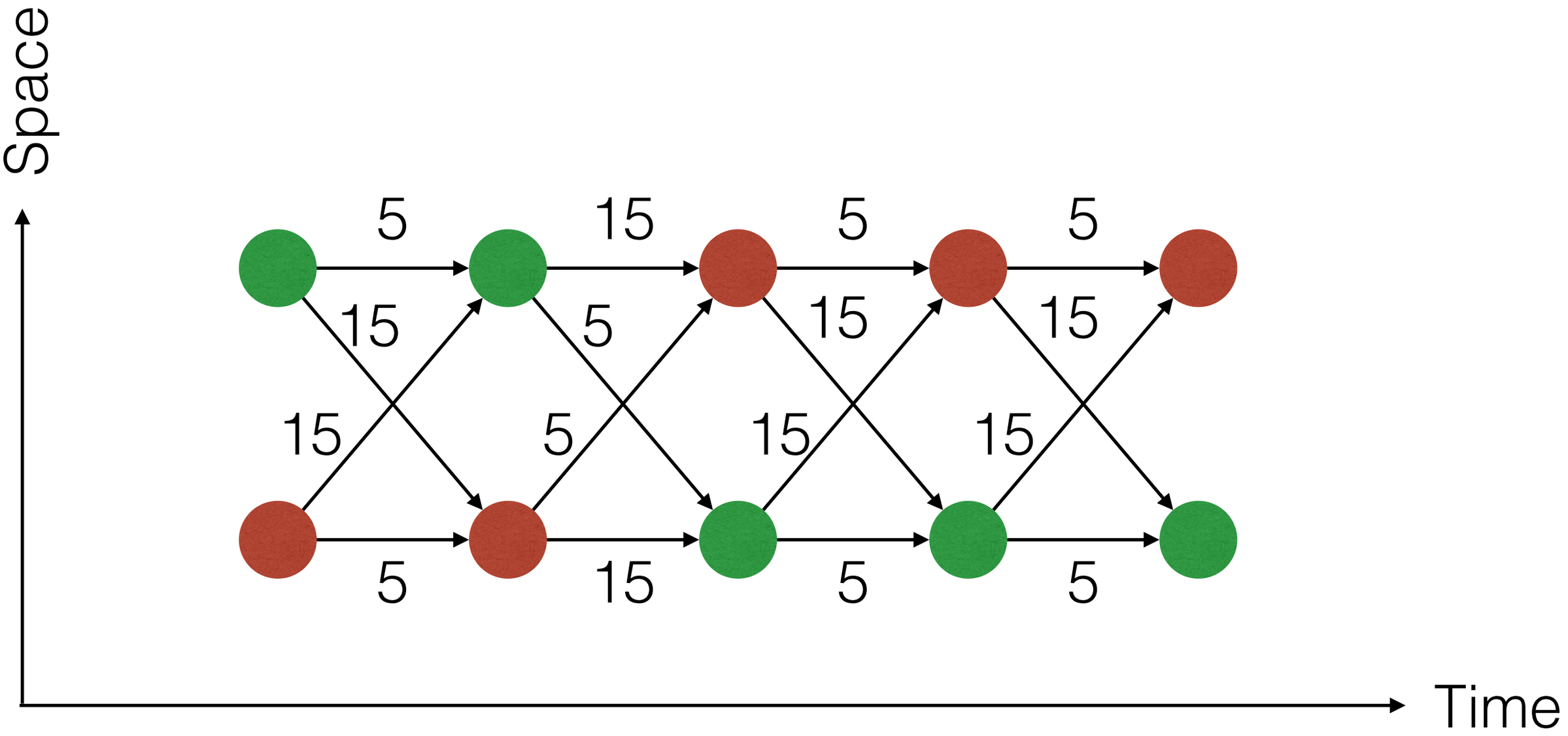
Edges are created between temporally close nodes.

Graph construction



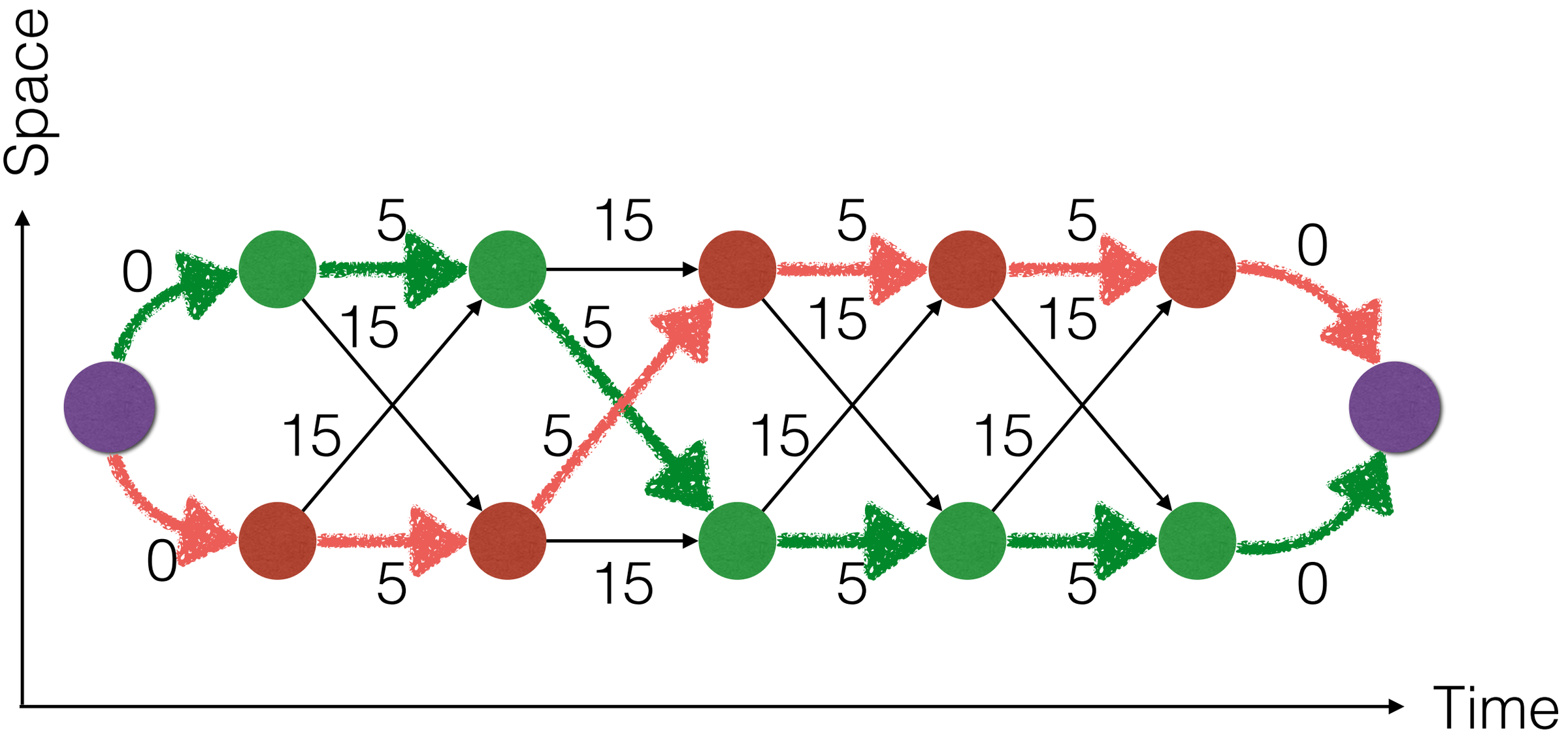
Edges are created between temporally close nodes.

Graph construction

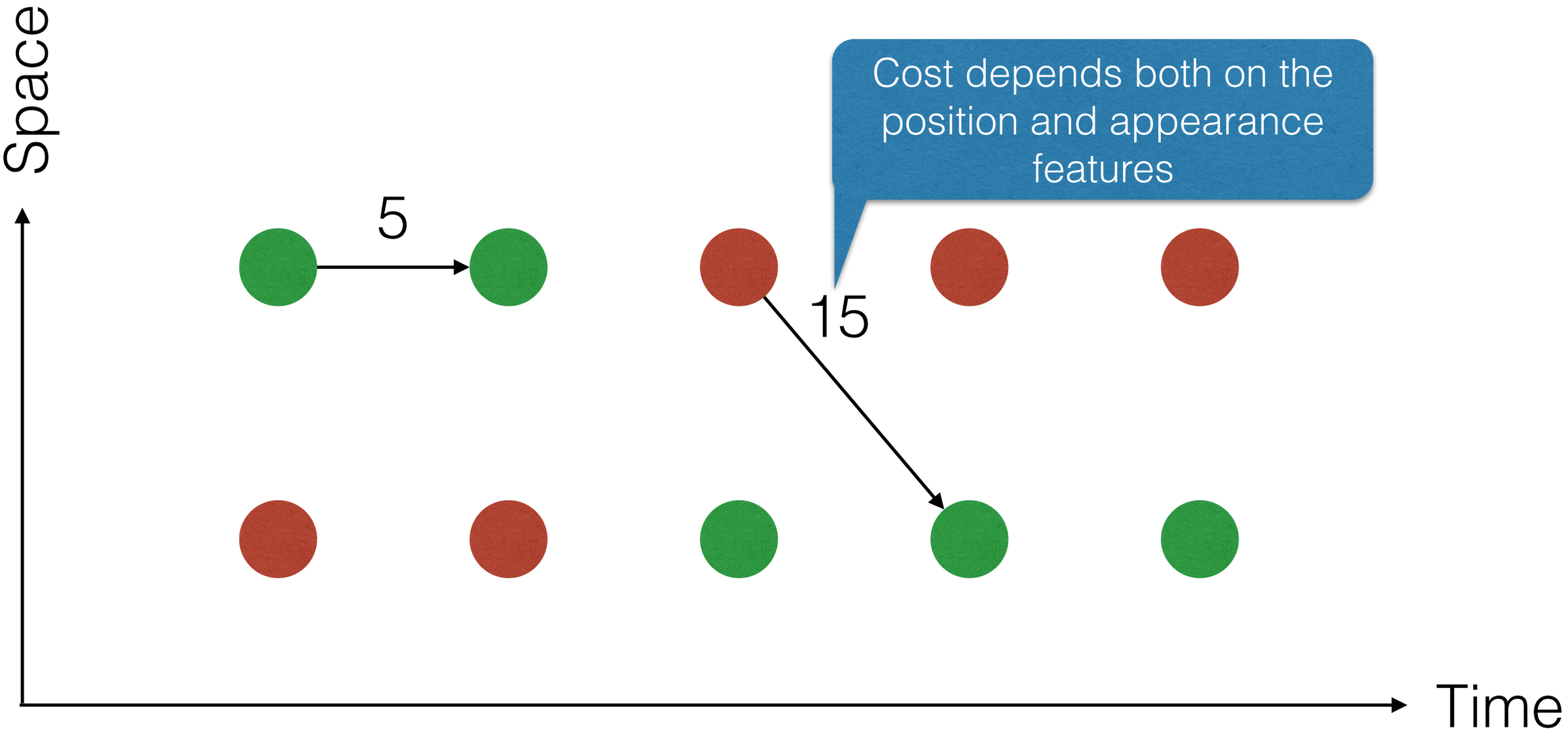


Edges are created between temporally close nodes.

Conventional K-shortest paths solution



Recall: graph construction



Edges are created between temporally close nodes.

What if…

- the appearance features are **not always reliable,**
- but, the **level of confidence can be predicted**?

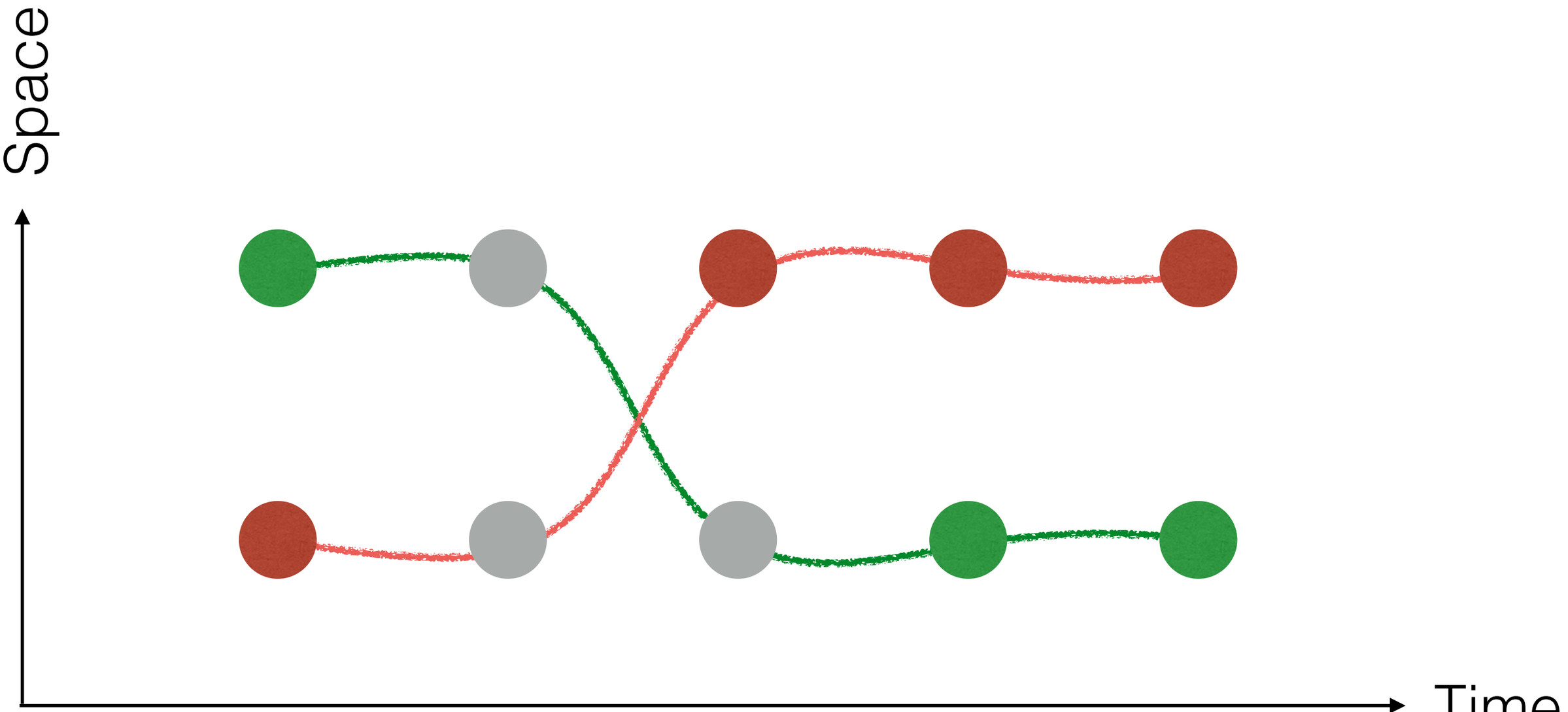
Facial recognition is available only when it faces the camera

Digit feature is available only when it faces the camera

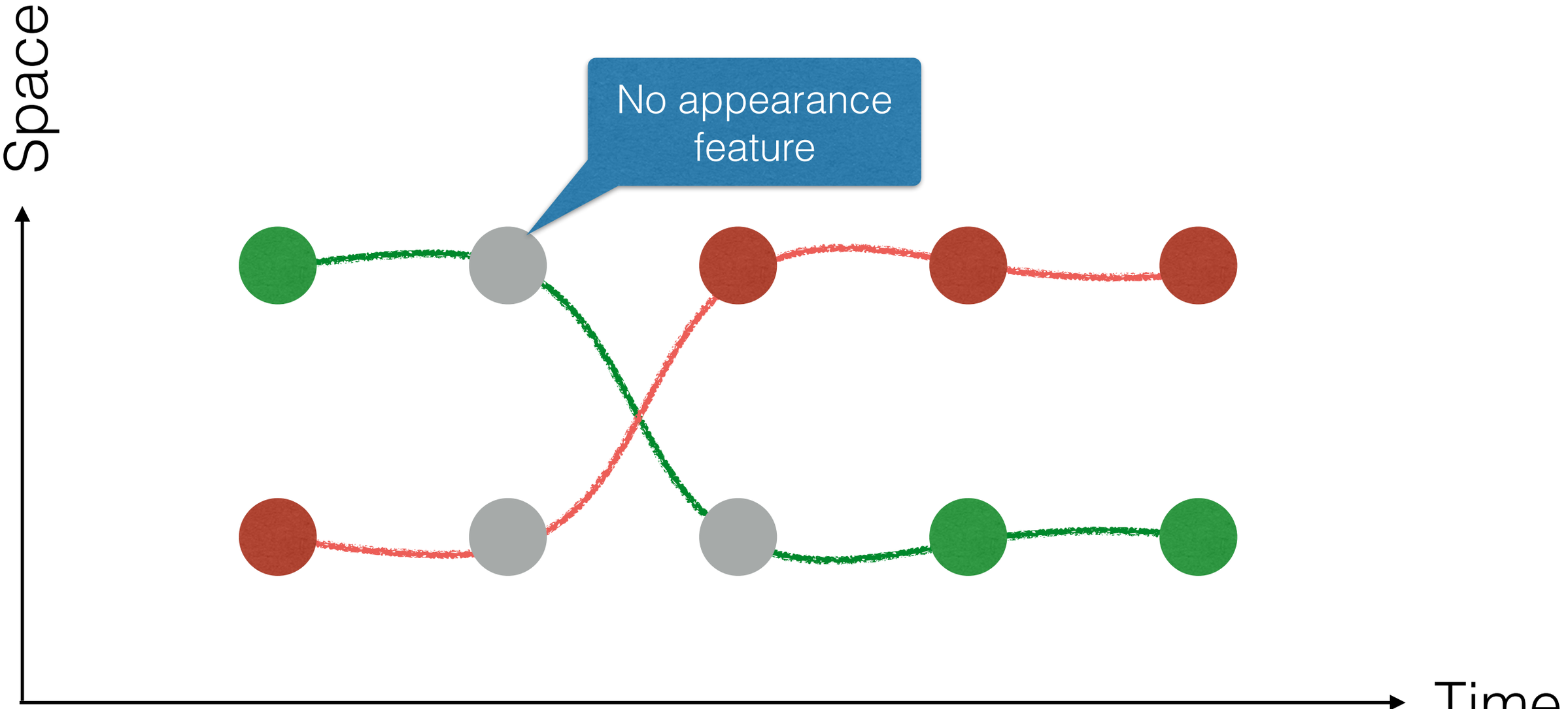
Color feature is noisy in presence of clutter, occlusion, etc.

Previous methods **cannot handle variable reliability in the features.**

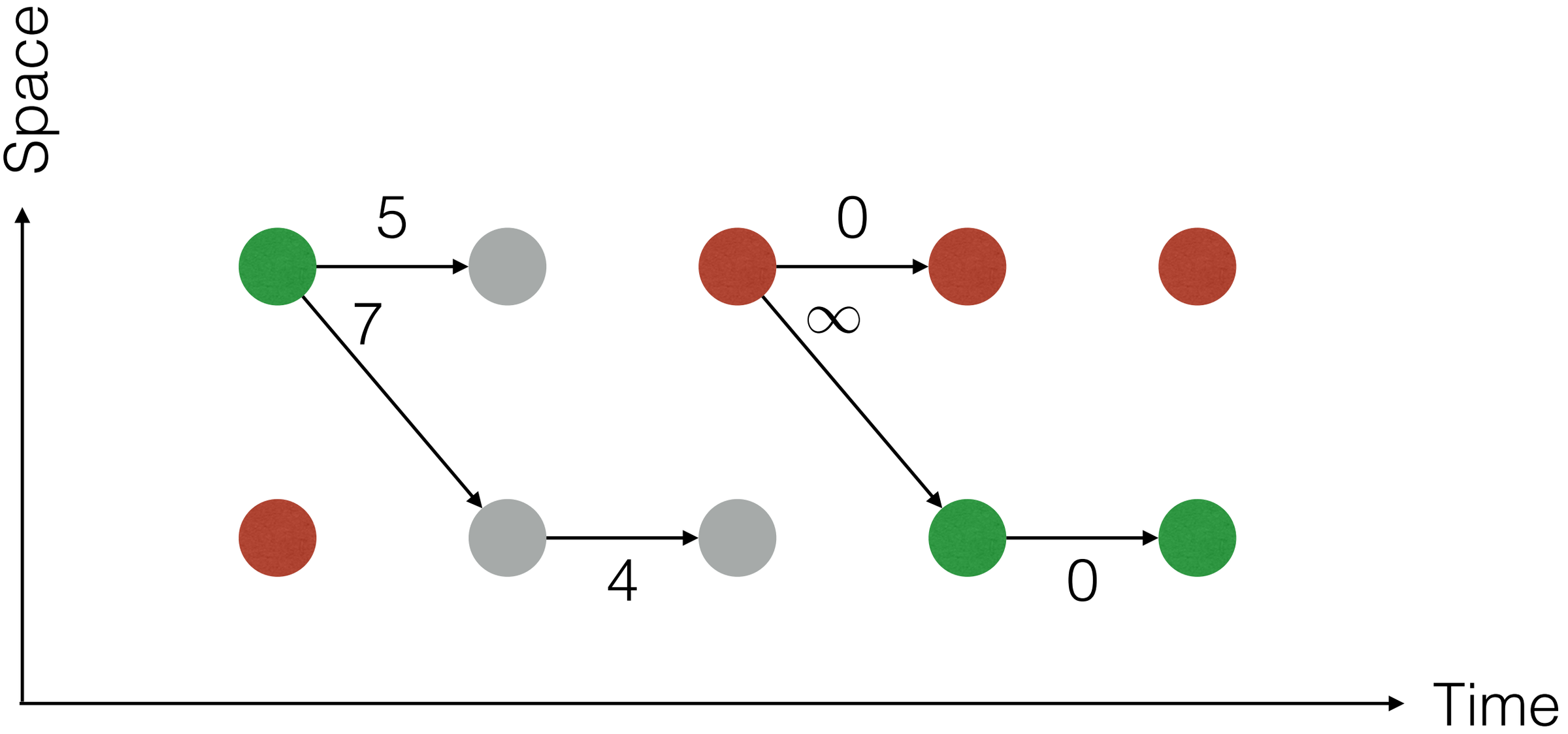
Detections



Detections

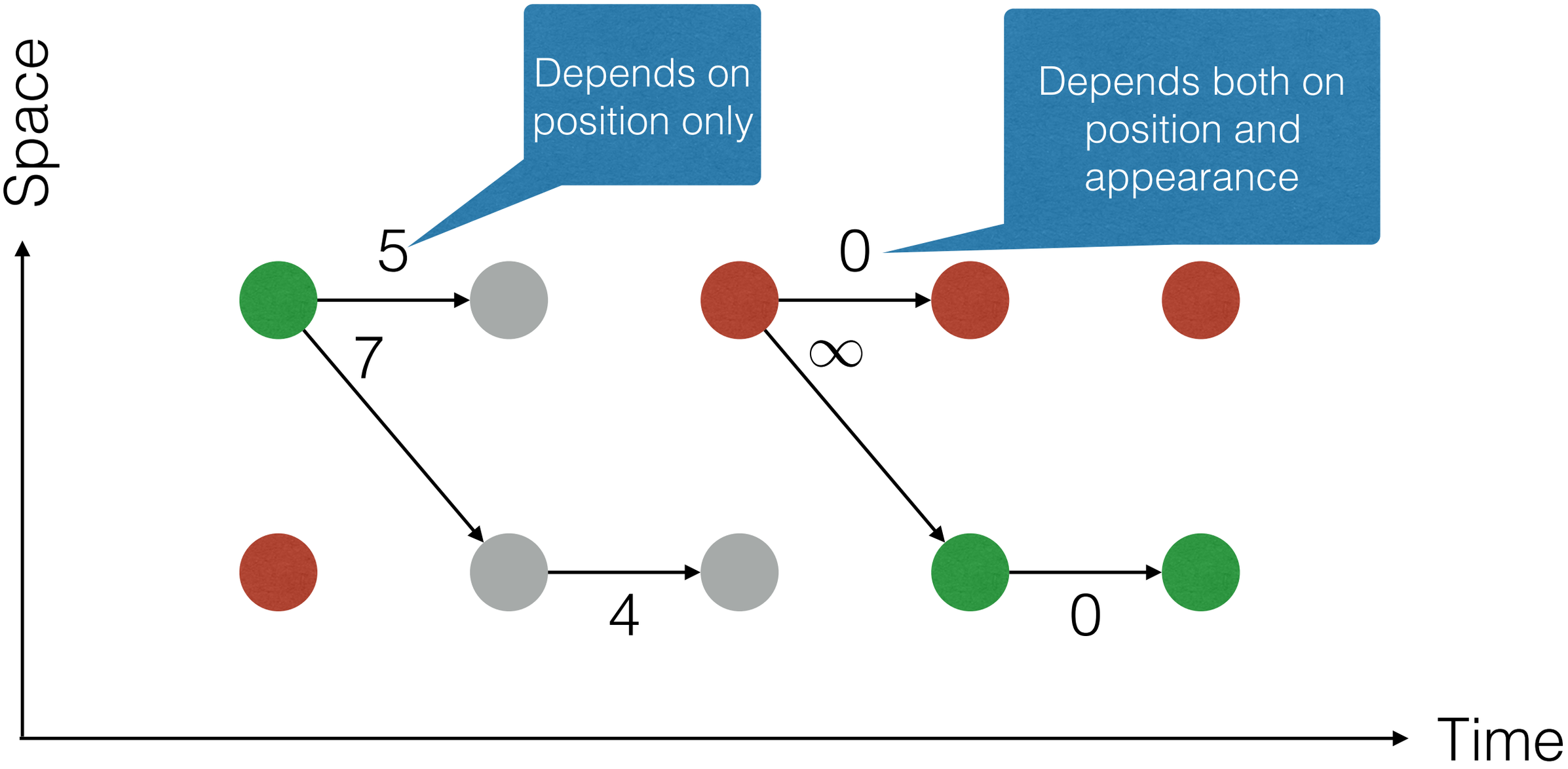


Graph construction



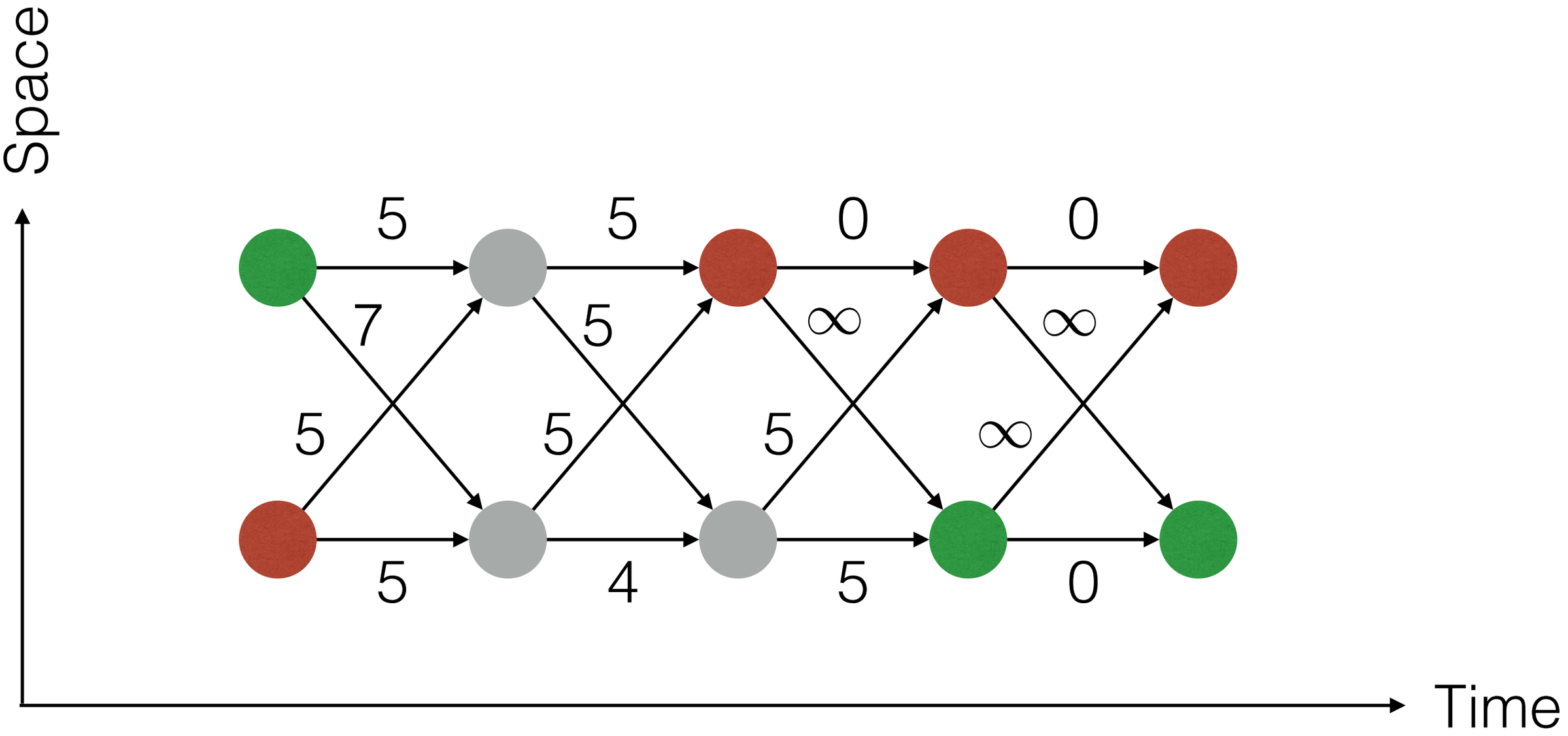
Edges are created between temporally close nodes.

Graph construction



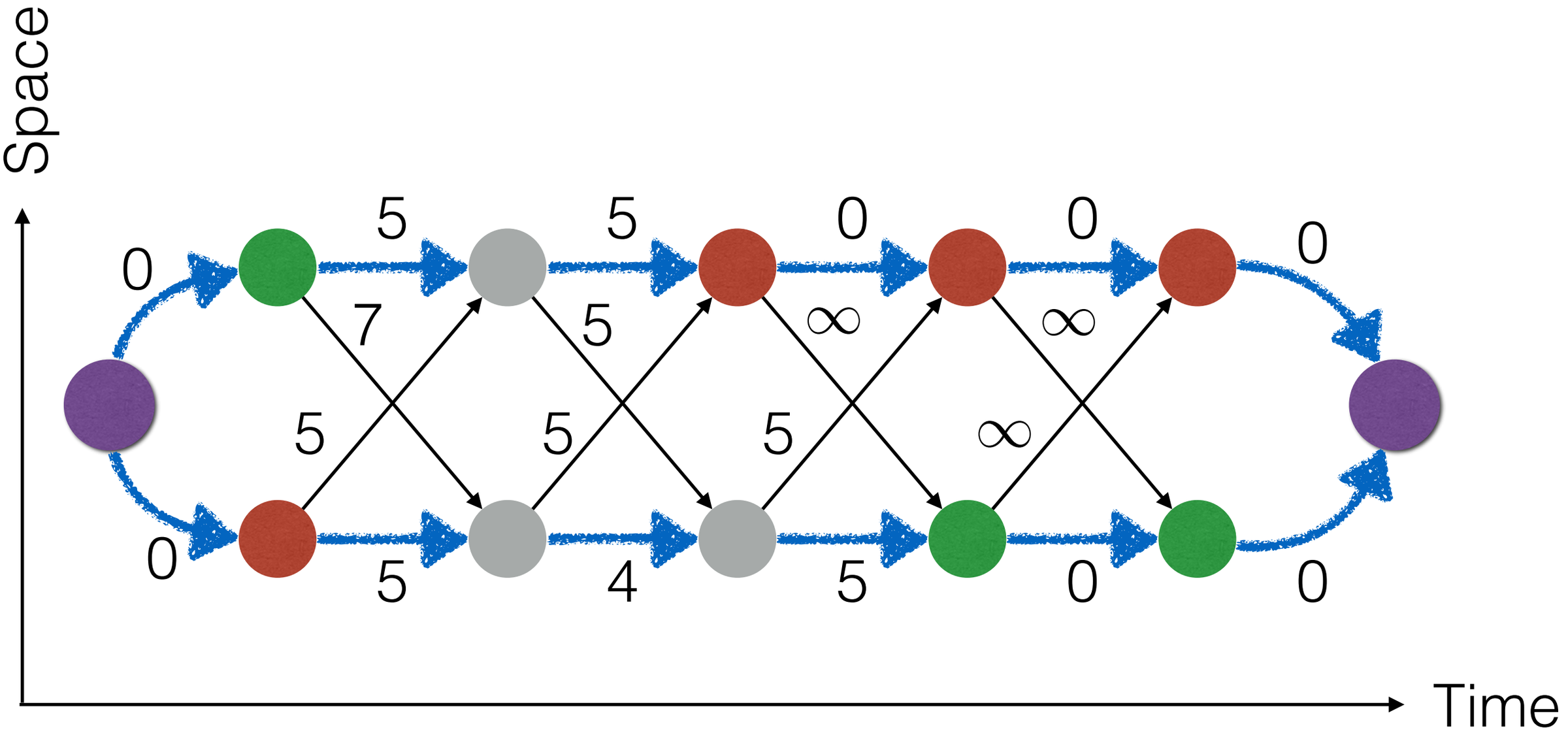
Edges are created between temporally close nodes.

Graph construction

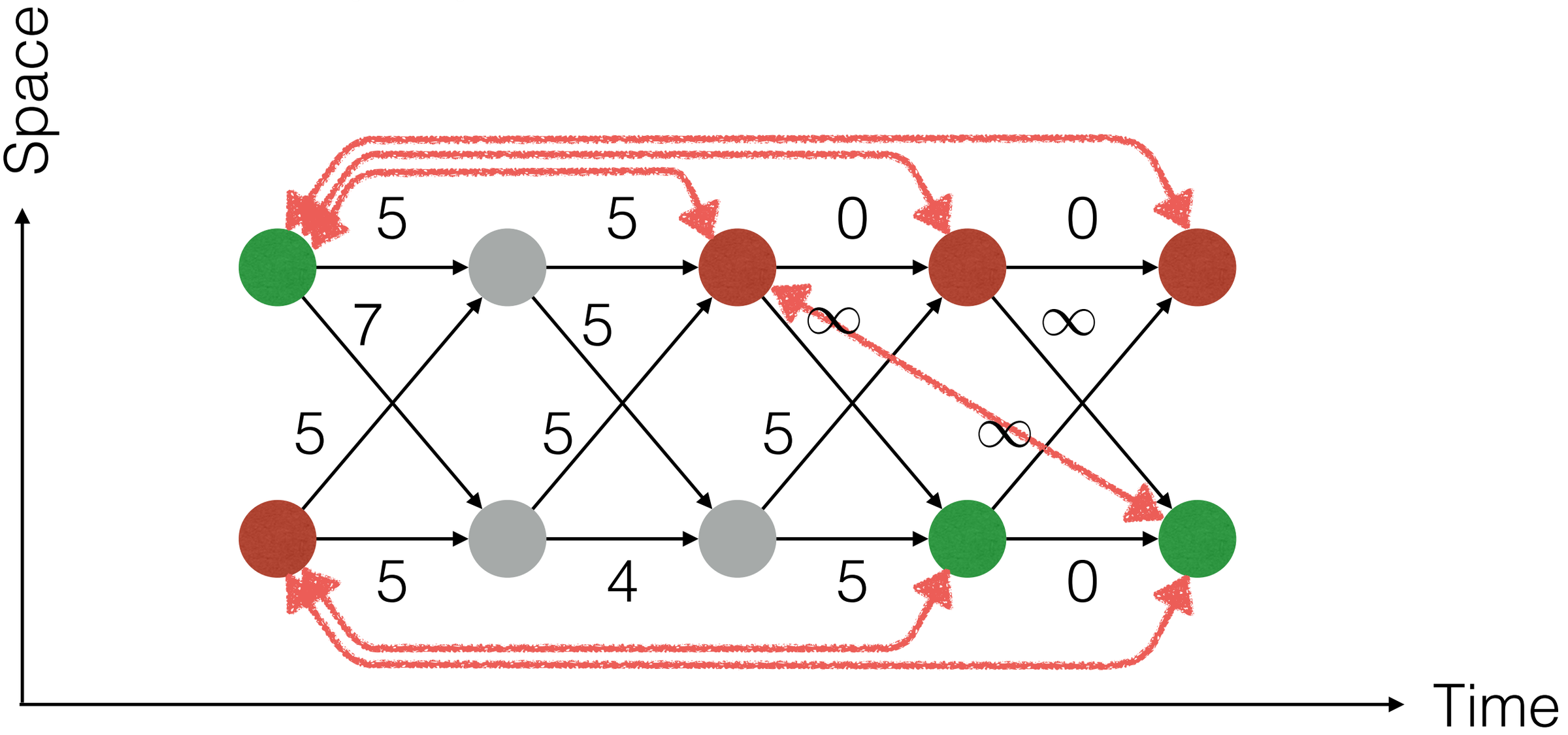


Edges are created between temporally close nodes.

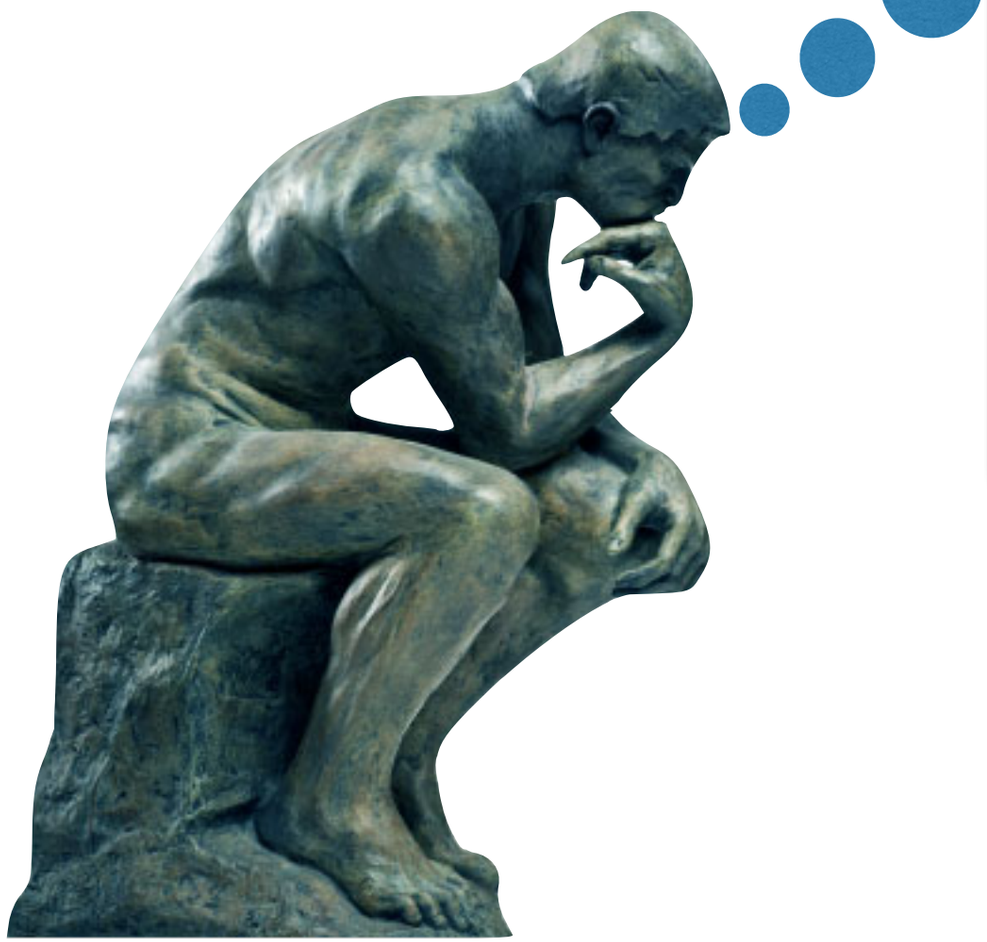
Conventional K-shortest paths solution



Cannot exploit appearance (in)consistencies.



Incompatibilities



How to design a multi-object tracker that handles such noisy/sporadic appearance features?

Contributions

- Iterative hypothesis testing *[ACCV 2012]*
- Prioritized belief propagation *[BMVC 2012]*
- Discriminative label propagation *[ICCV 2013, TPAMI 2015 (under revision)]*

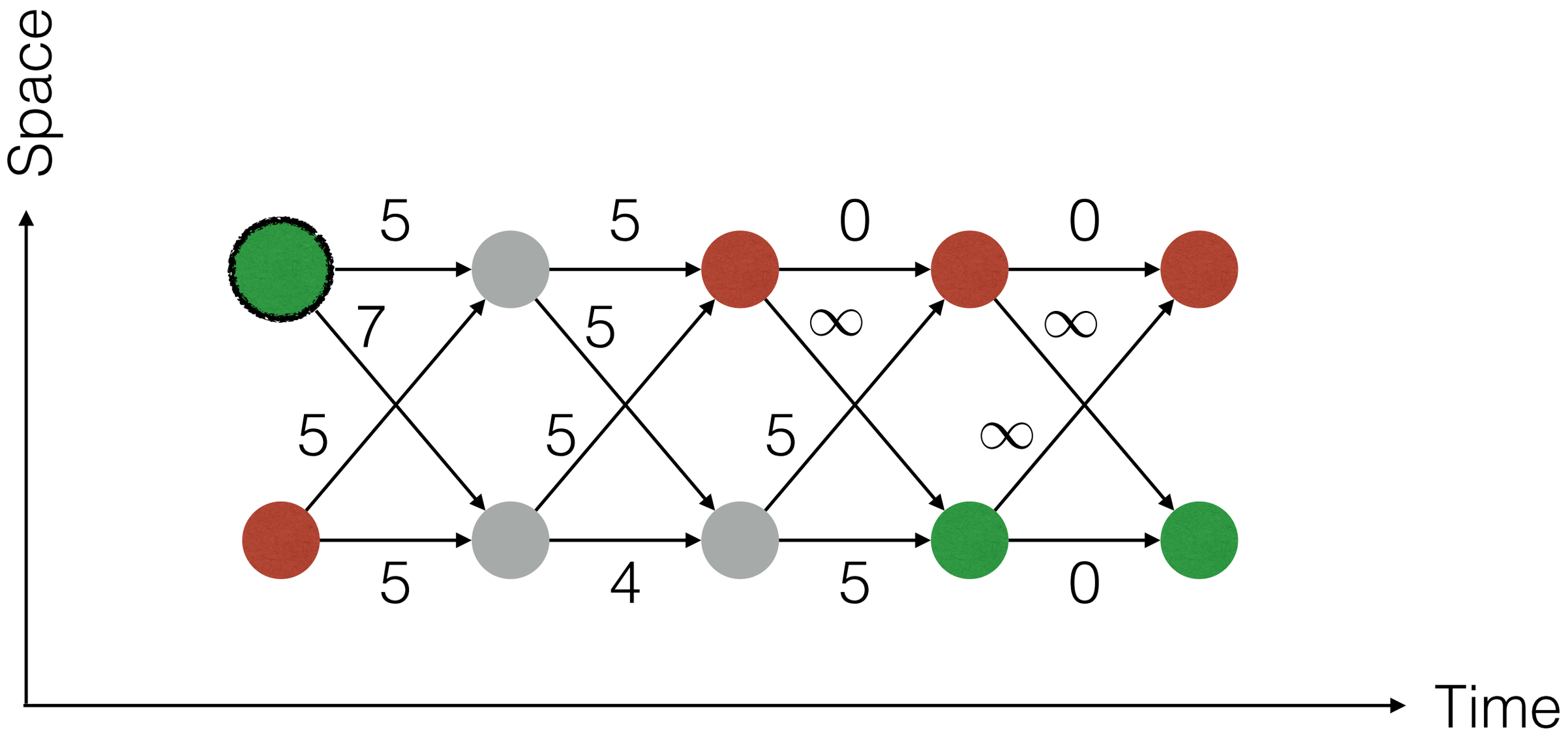
Contribution 1

Iterative hypothesis testing (IHT)

ACCV 2012

Jointly with Damien Delannay

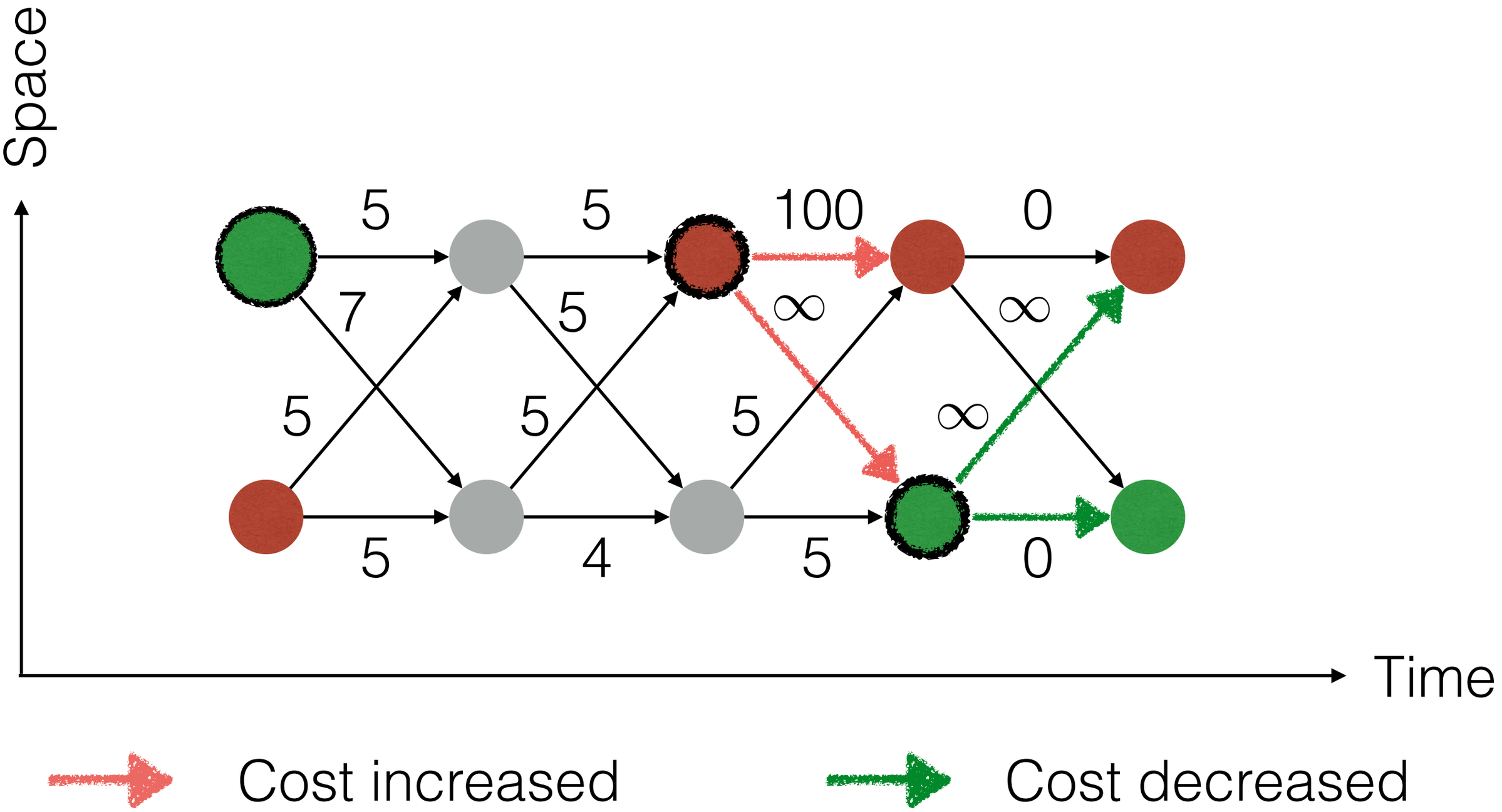
Make an assumption about target appearance (1)



Contribution 1

Hypothesis:

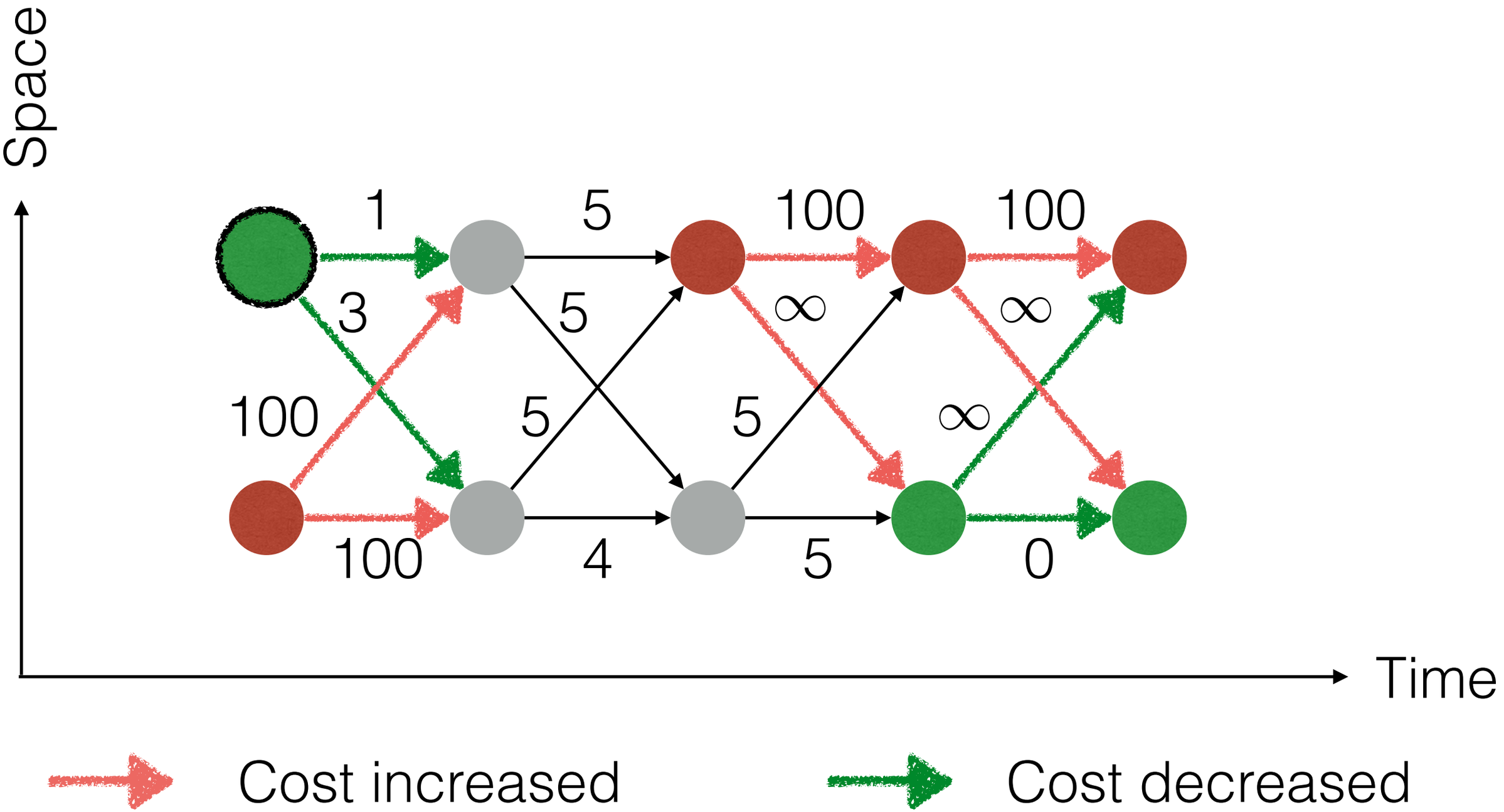
Key-node appearance=target appearance



Contribution 1

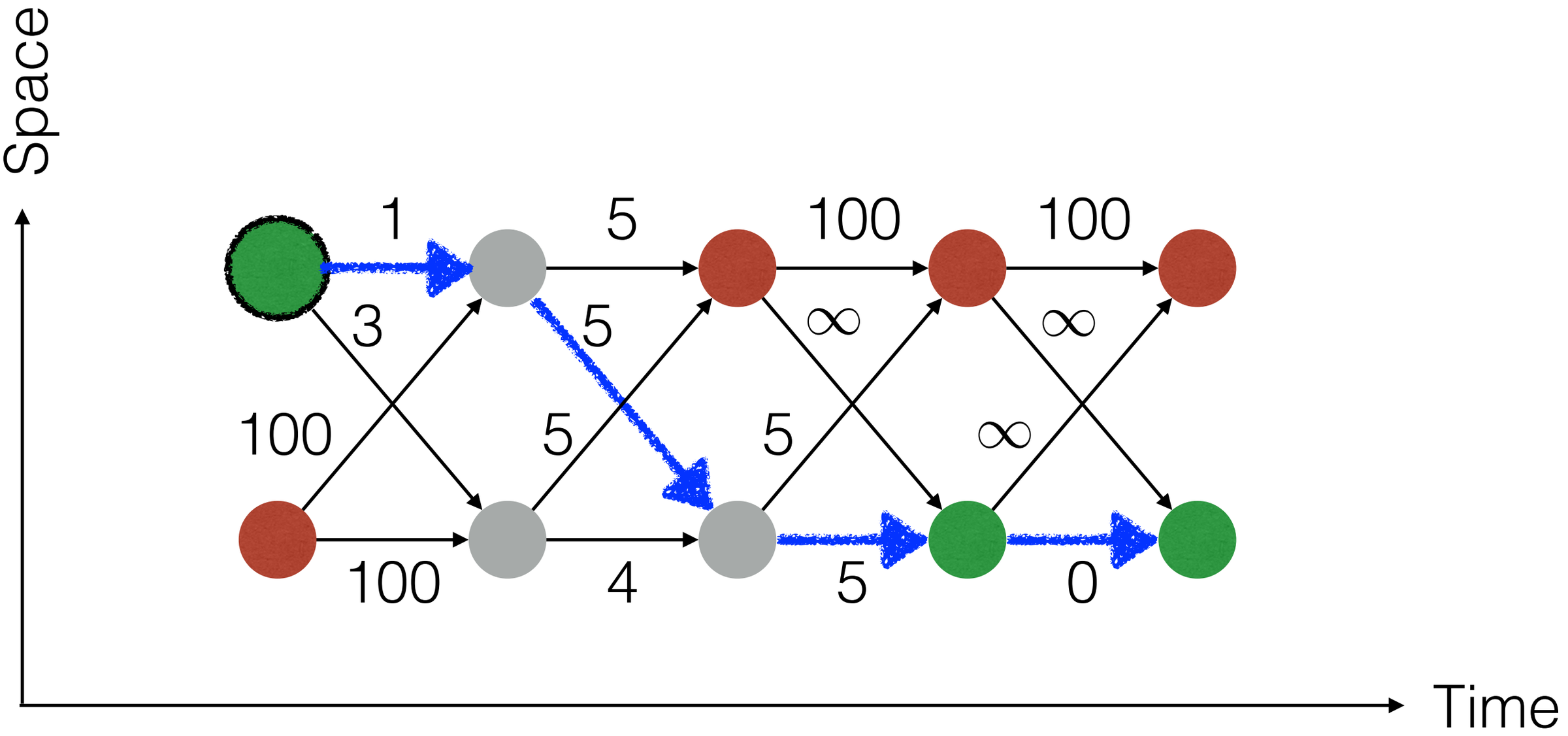
Hypothesis:

Key-node appearance=target appearance



Contribution 1

Compute shortest-path

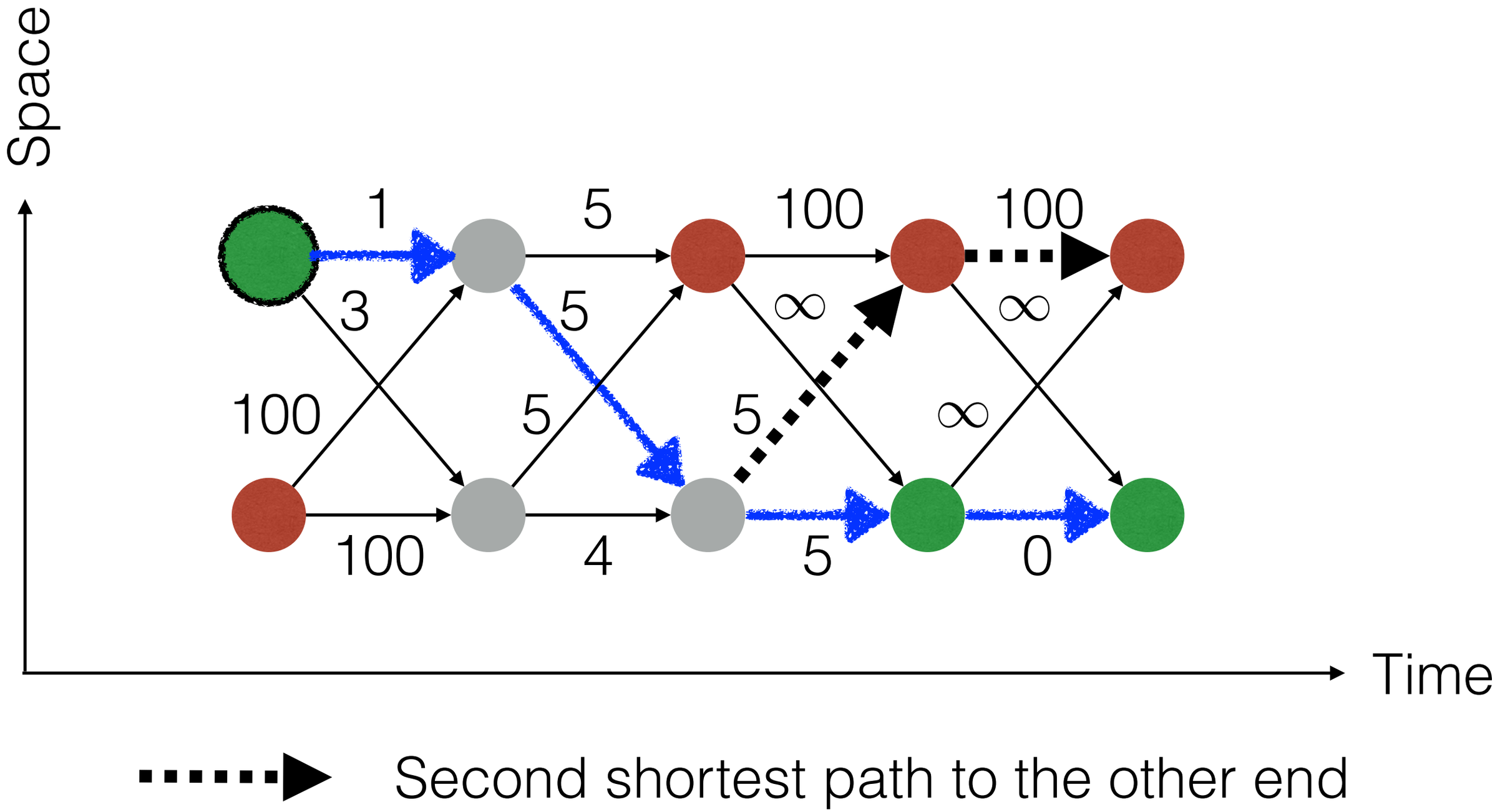


Contribution 1

Issues

- Which appearance first?
  - **Iterative**
- How to check if the shortest path is good?
  - **Hypothesis testing**

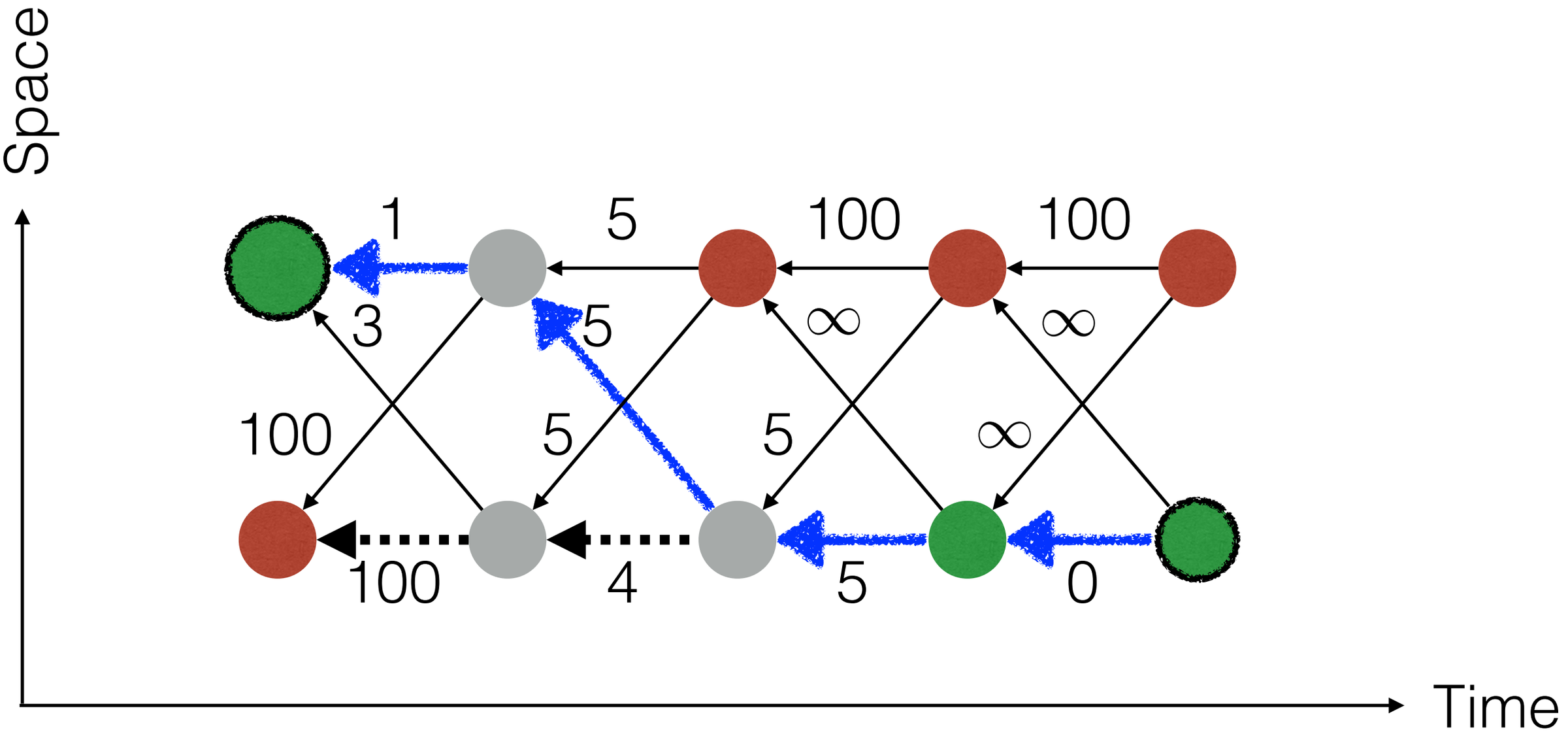
Test the shortest-path (1)



Second shortest path to the other end

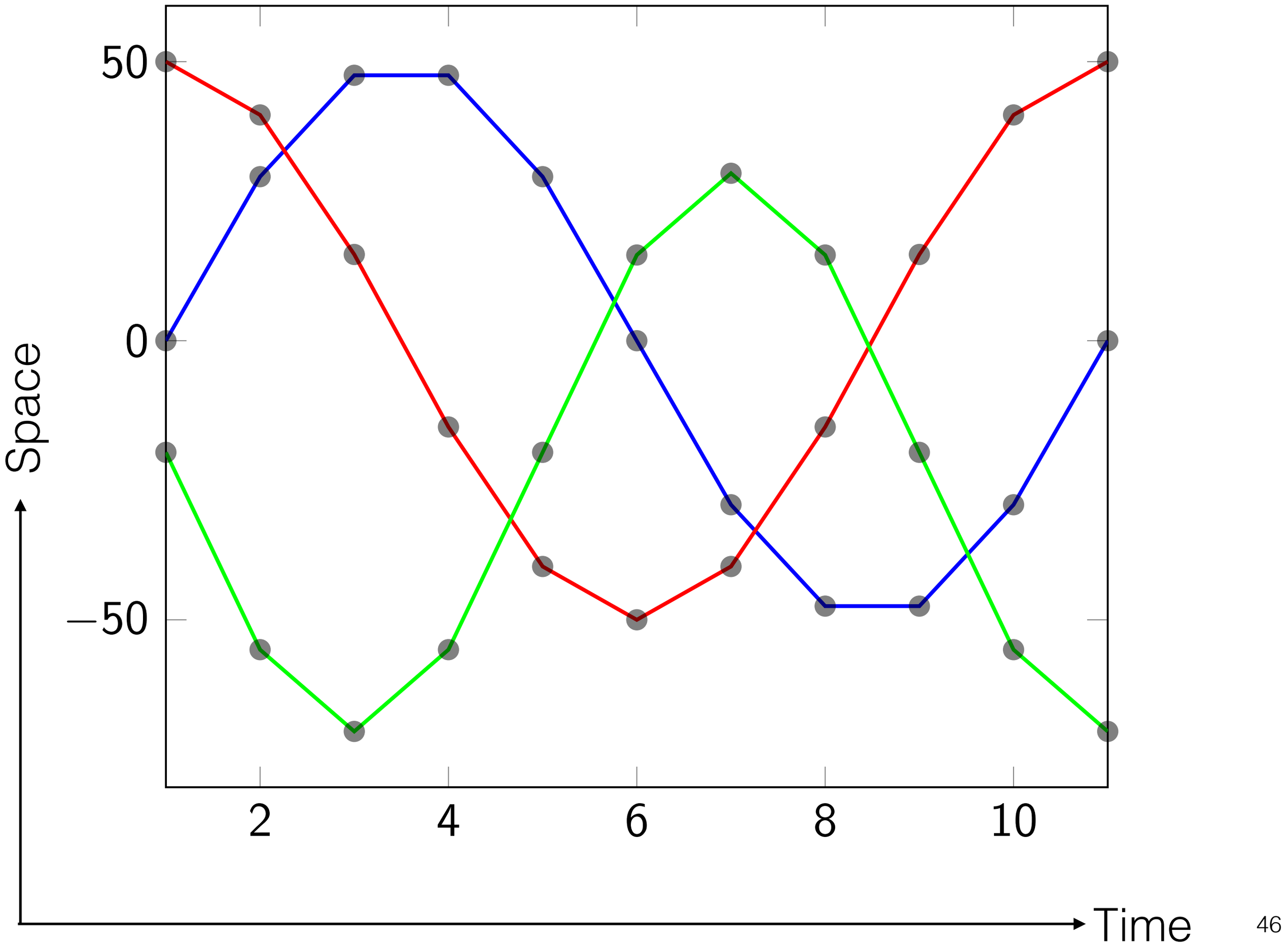
Contribution 1

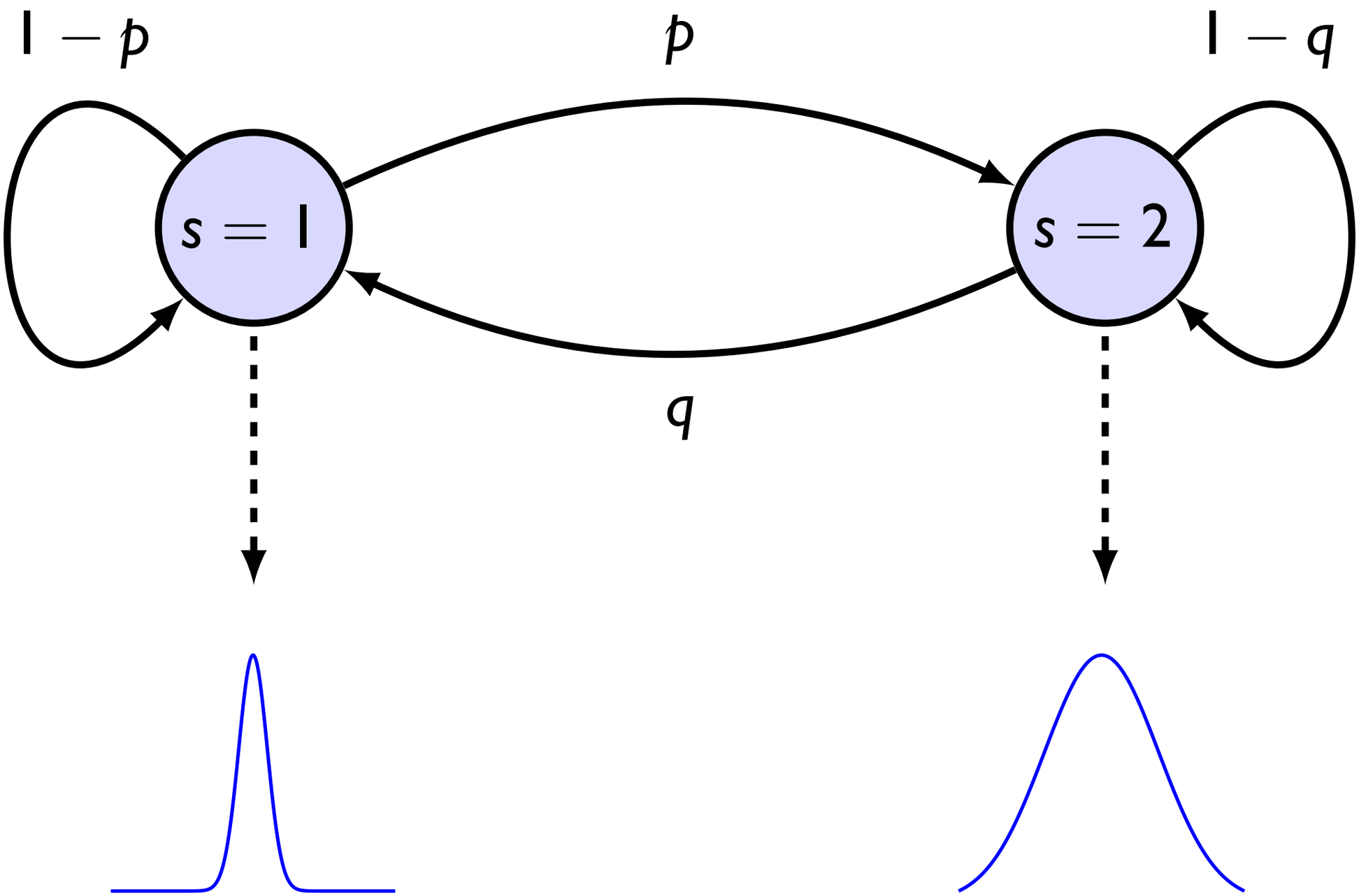
Test the shortest-path (2)

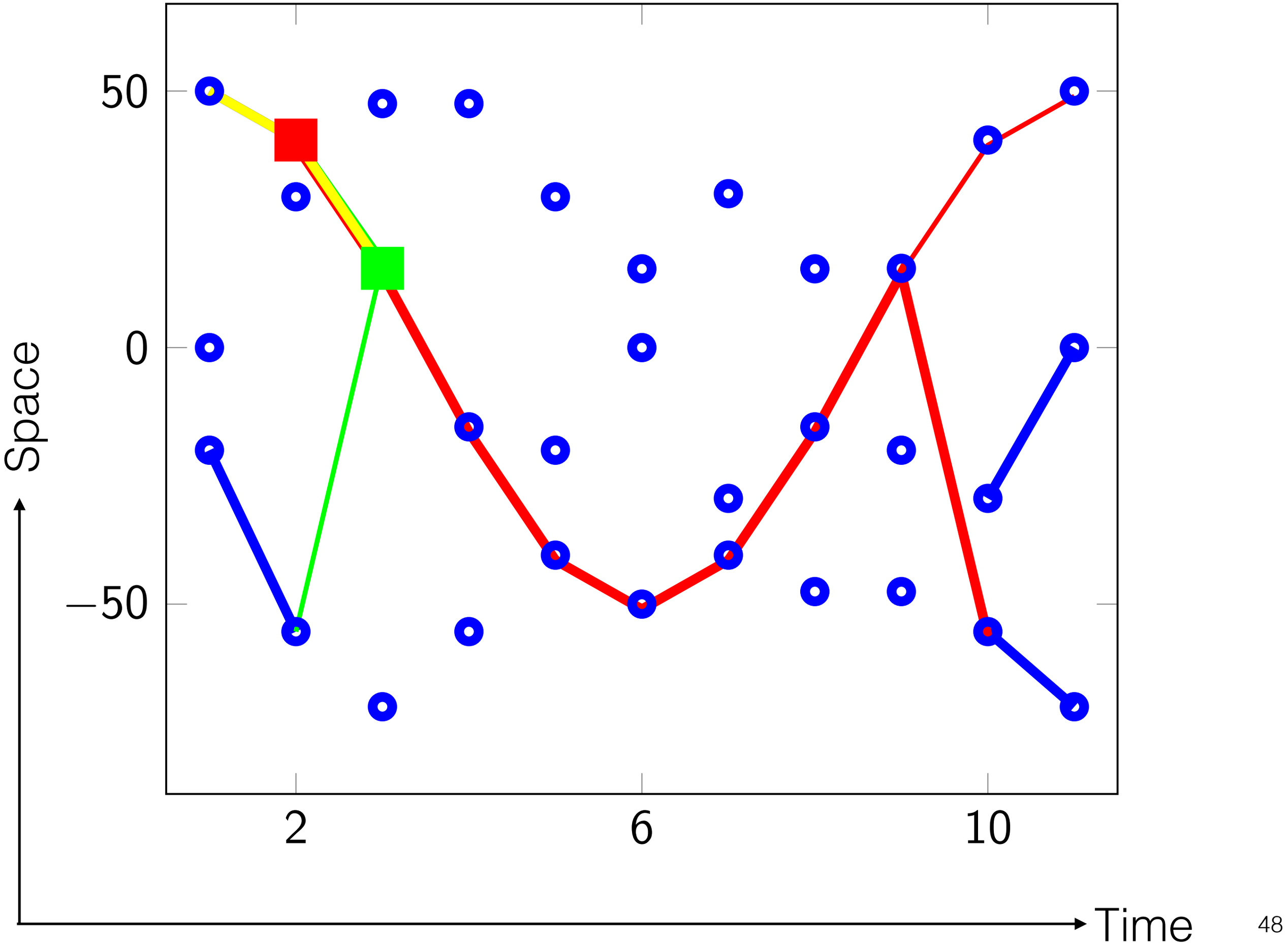


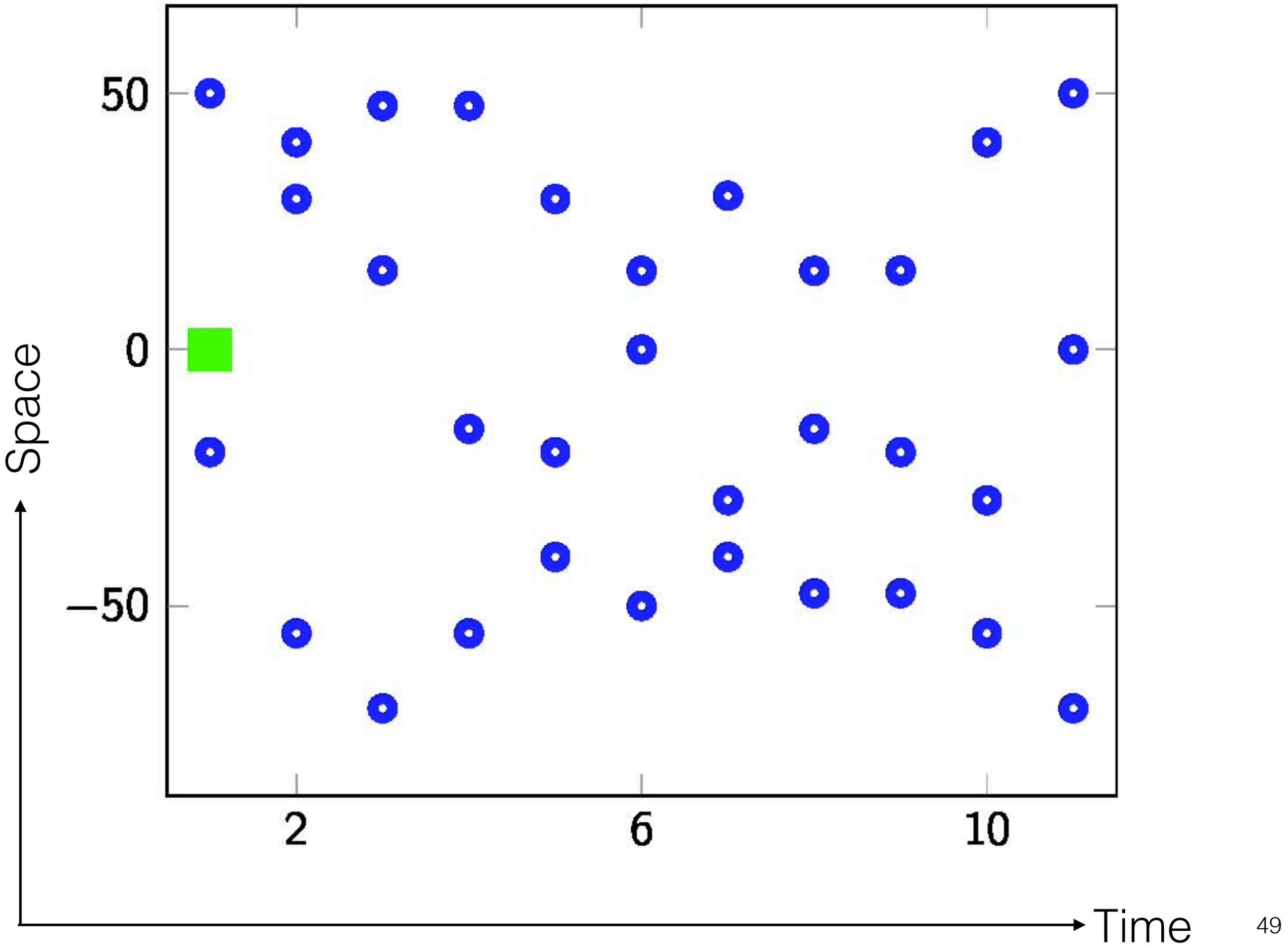
Contribution 1

Demo


50
0
−50
Space
2
4
6
8
10
Time


1 − p
p
1 − q
s = 1
s = 2
q

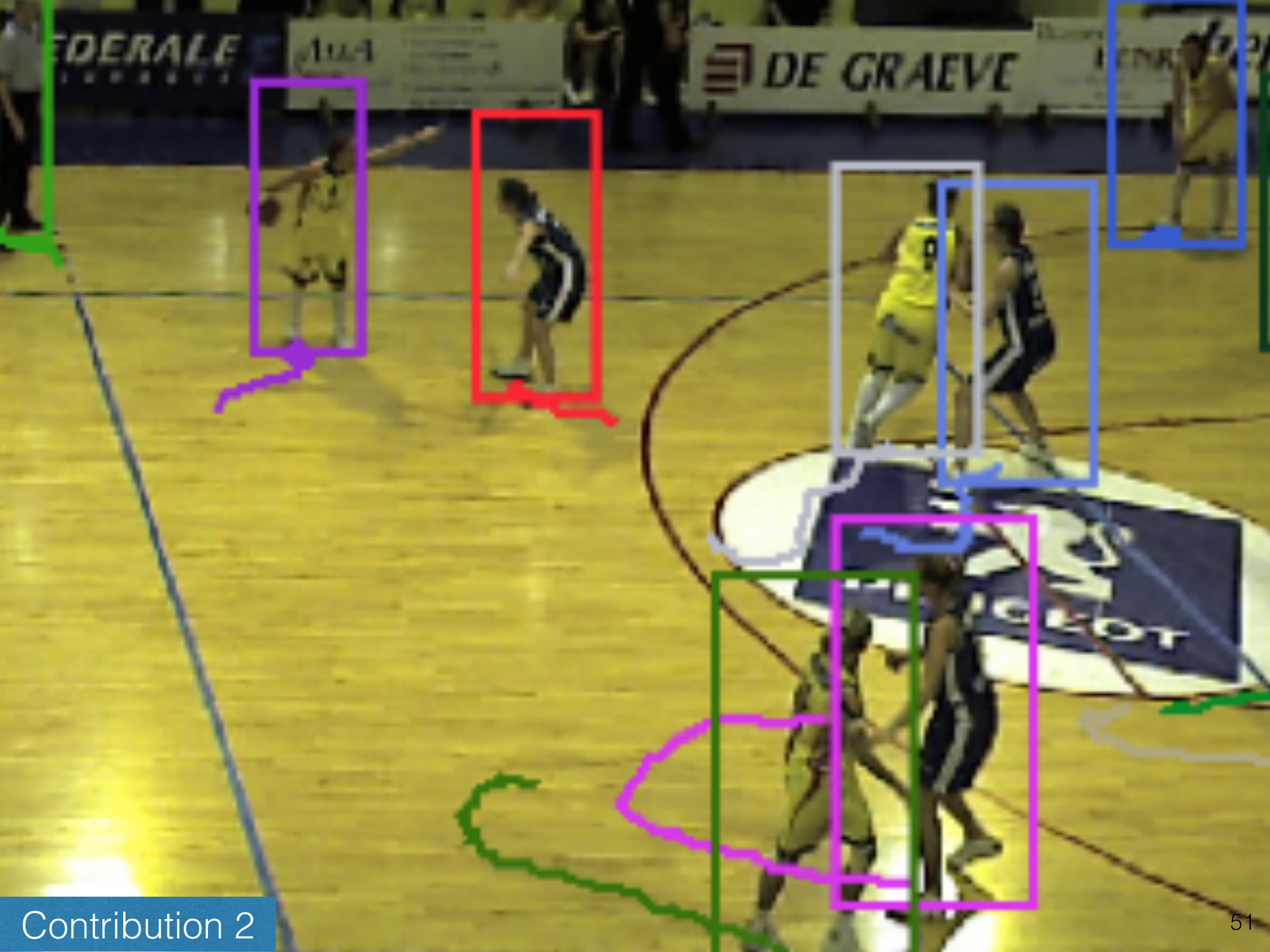

50
0
−50
Space
2
6
10
Time


50
0
−50
Space
2
6
10
Time

Contribution 2

Prioritized belief propagation (PBP)

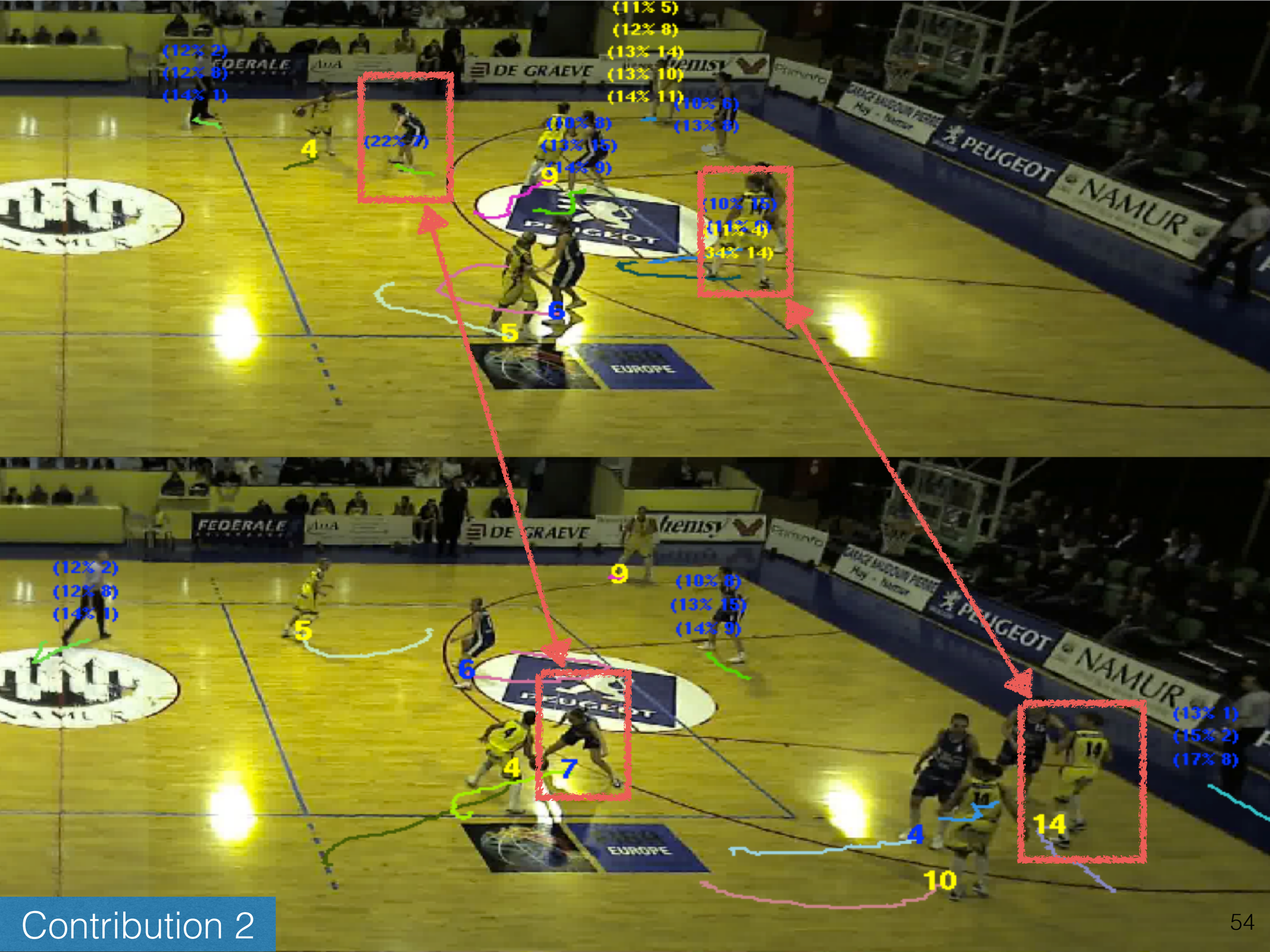
BMVC 2012



Contribution 2

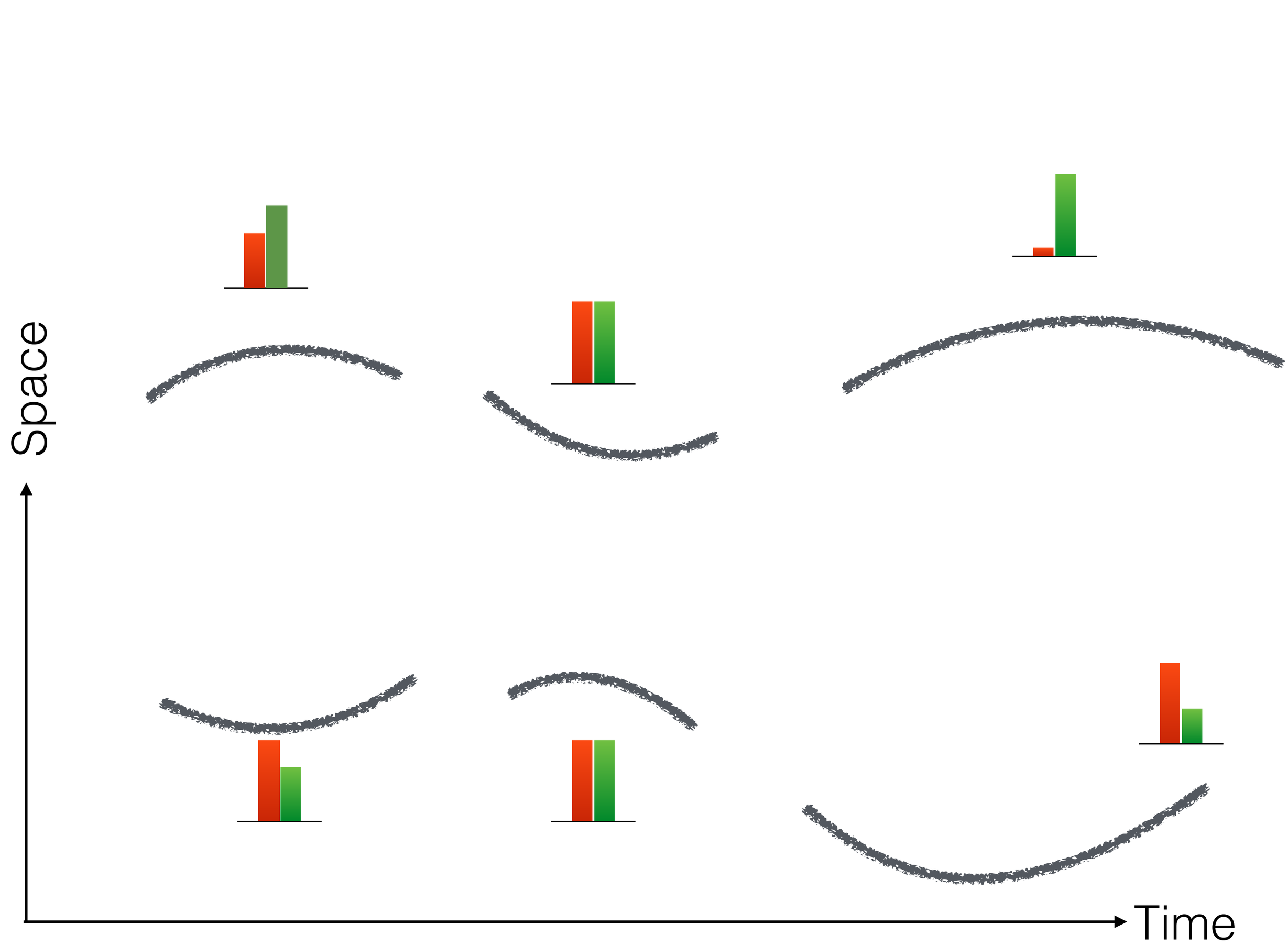

(19% 8)
(21% 10)
(22% 11)
4
7
9
9
6
5

Who is who?



Contribution 2

Contribution 2



Contribution 2

Space

Less ambiguous

More ambiguous

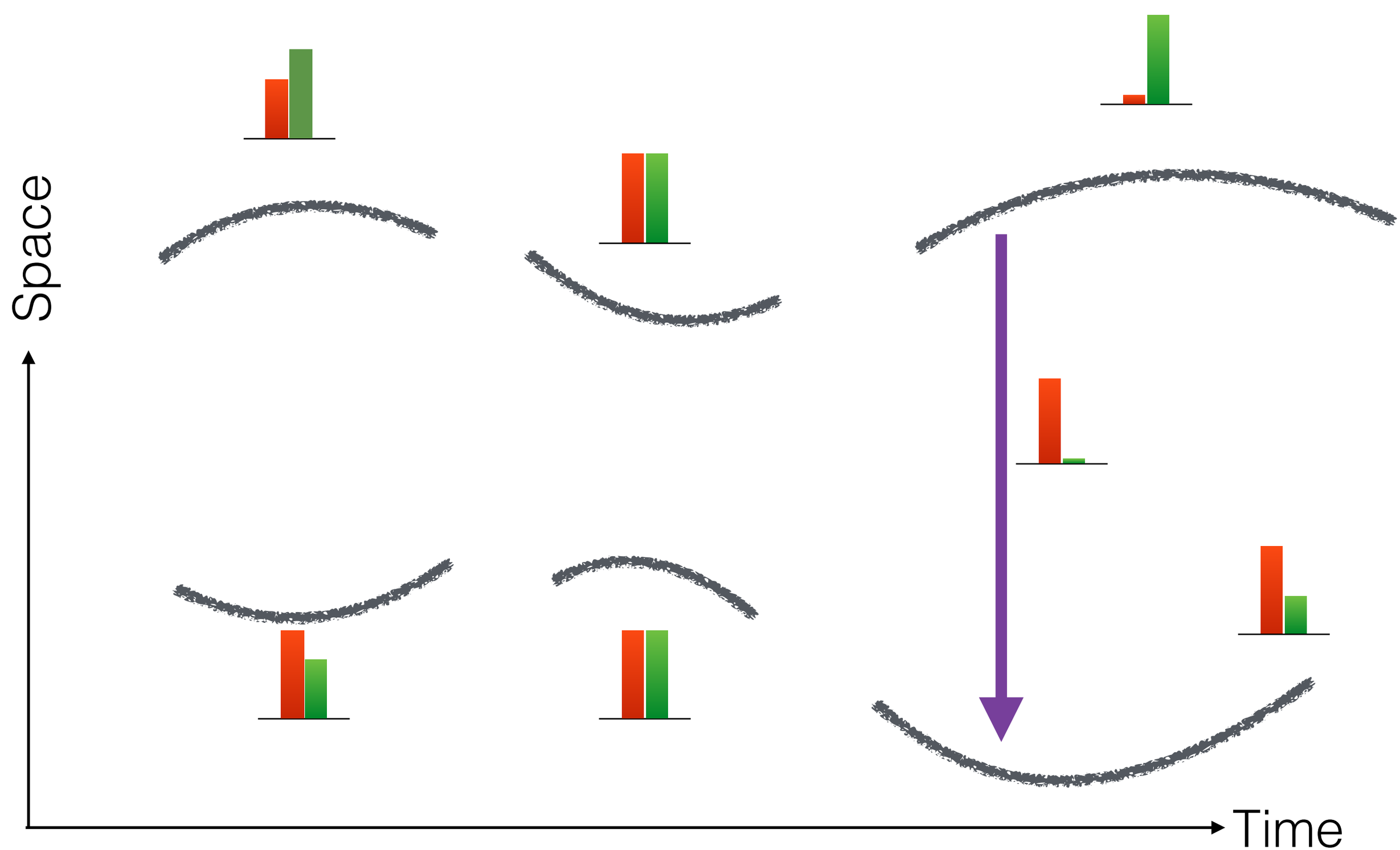
Time

Contribution 2

Message passing

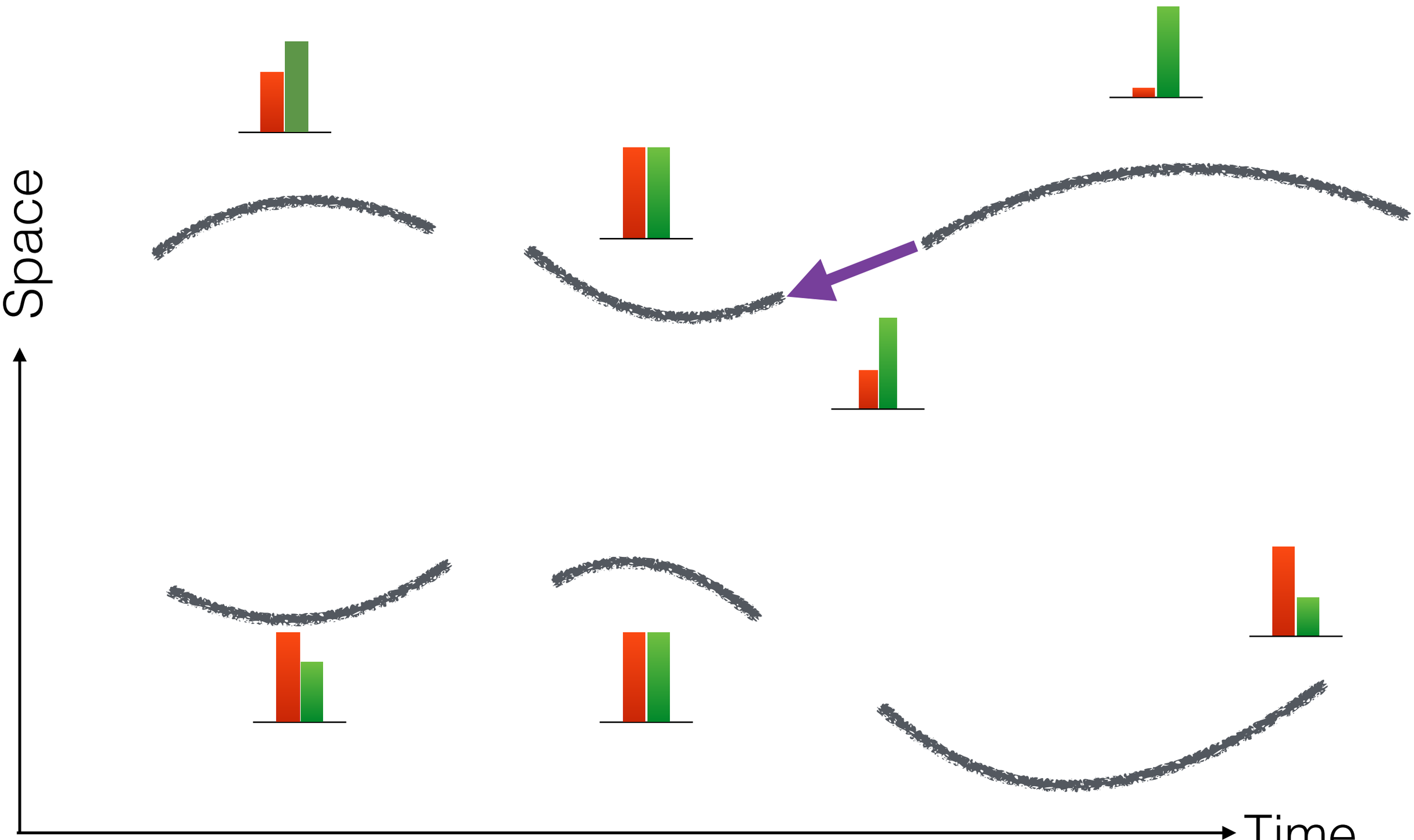
Contribution 2

You CANNOT be me.



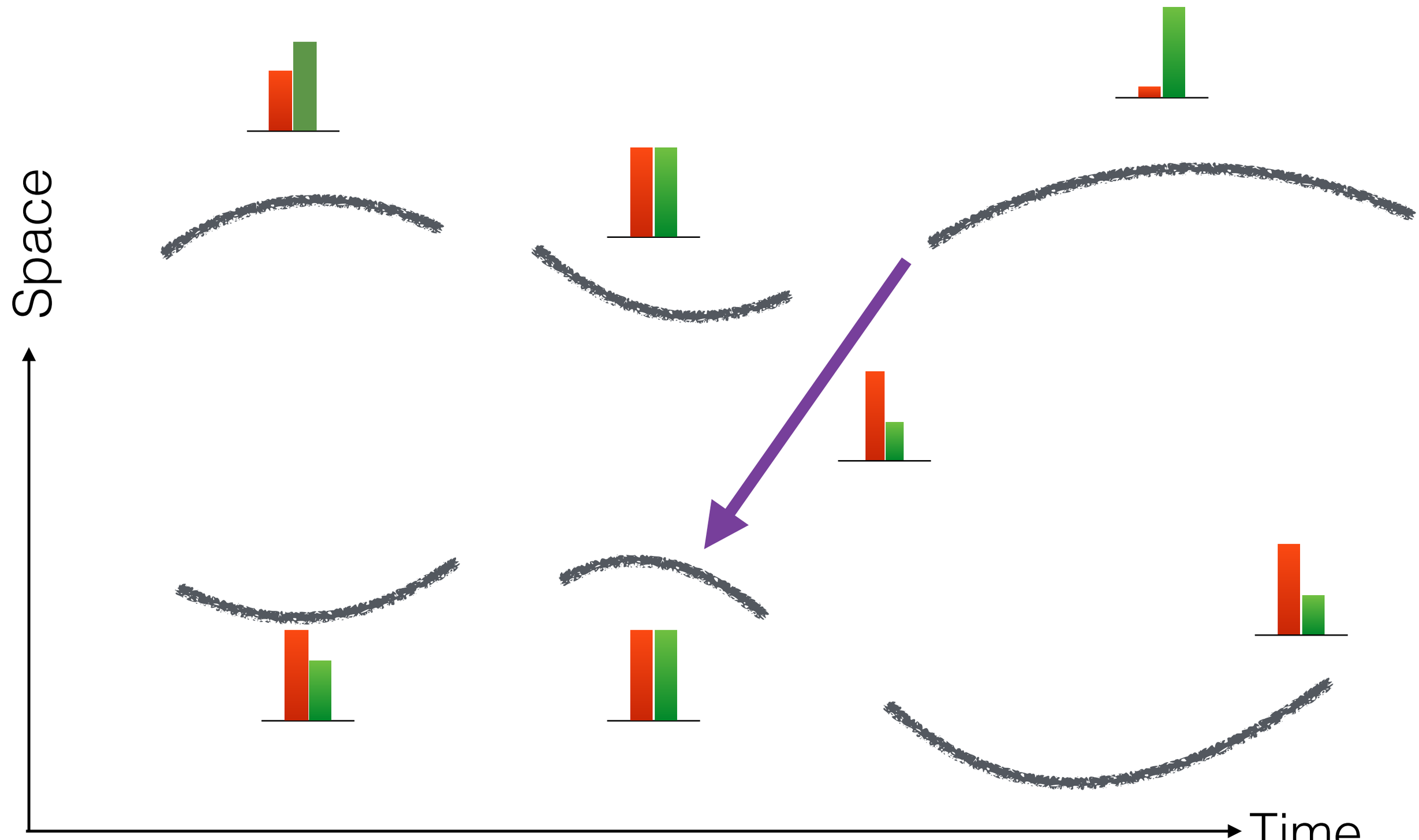
Contribution 2

You look MORE like me.



Contribution 2

You look LESS like me.

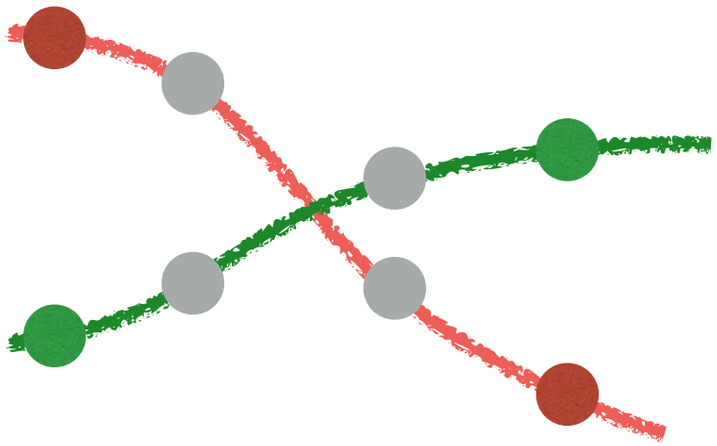


Contribution 2

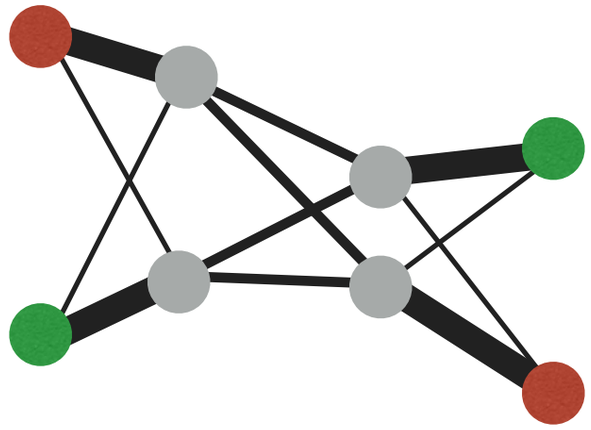
Contribution 3

Discriminative label propagation (DLP)

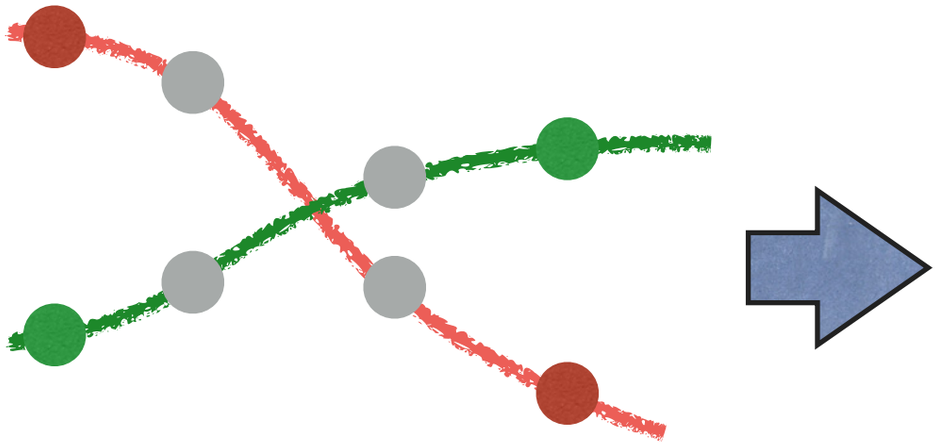
ICCV 2013, TPAMI 2015 (under revision)



Two targets with detections

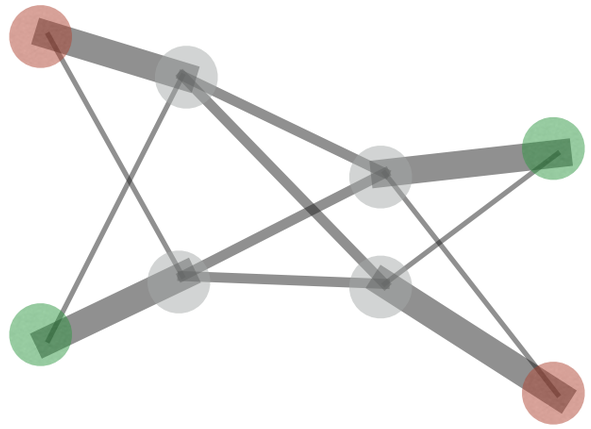


<u>Spatio-temporal graph:</u>
small displacement
≈
same target

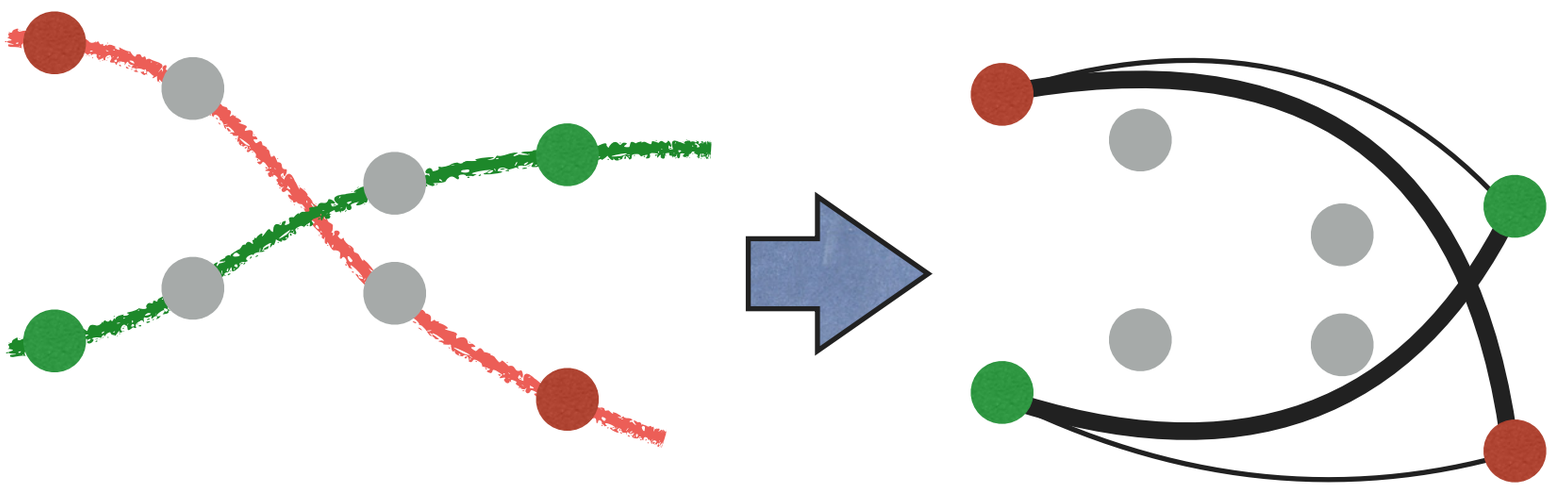


Two targets with detections

Contribution 3

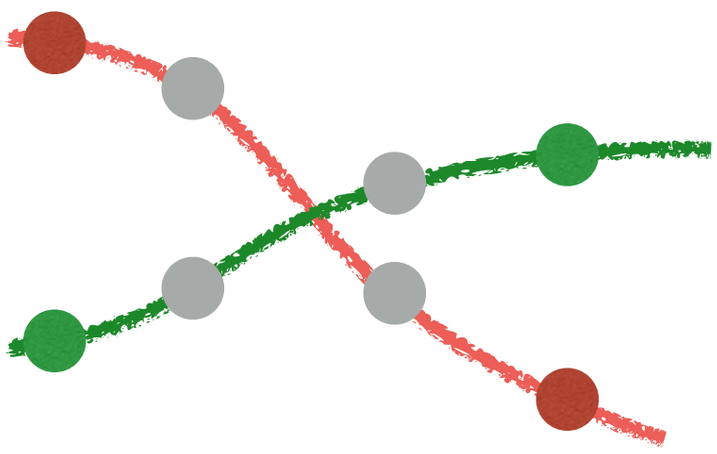


Spatio-temporal graph:
small displacement
≈
same target

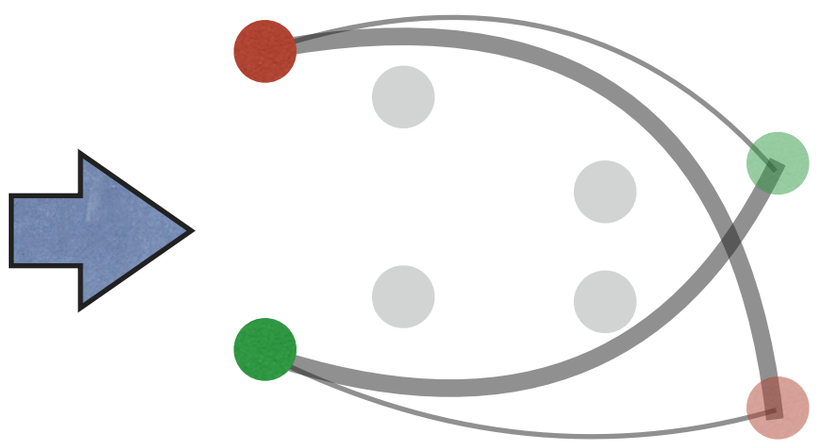


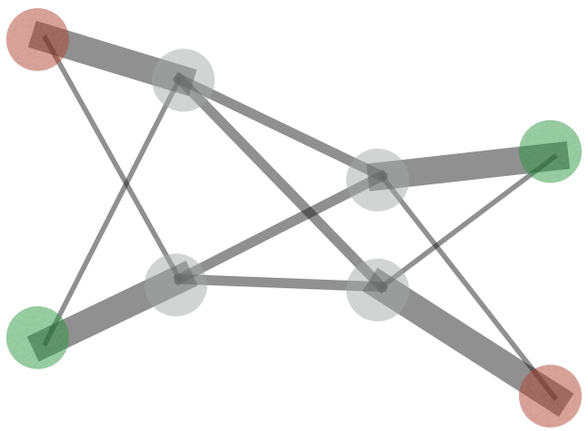
Two targets with detections

Appearance graph:
same appearance
≈
same target



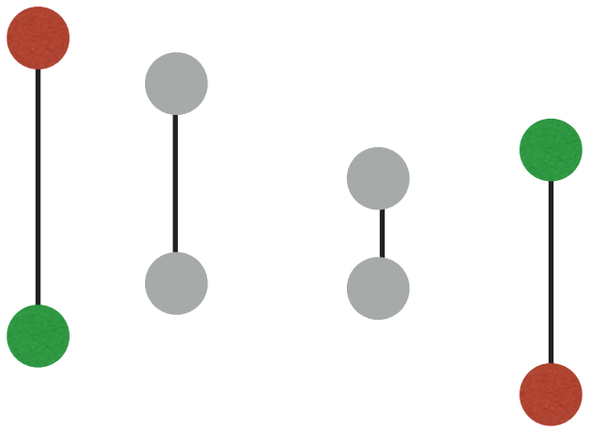
Two targets with detections





Spatio-temporal graph:
small displacement
≈
same target

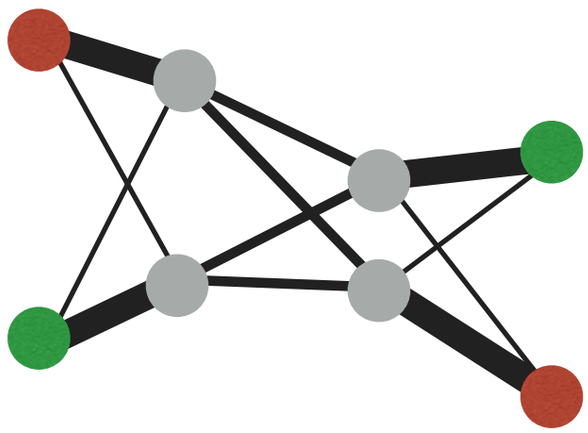
Appearance graph:
same appearance
≈
same target



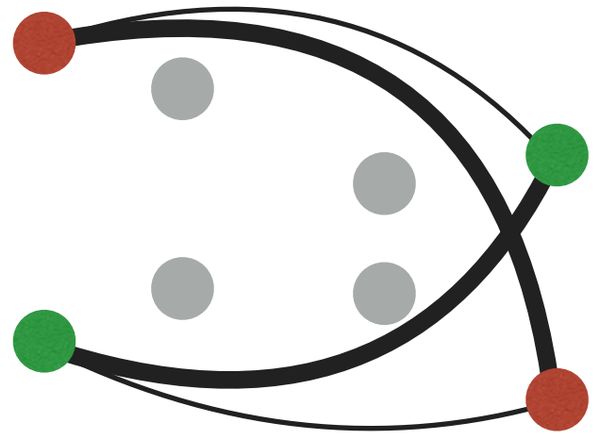
Exclusion graph:
co-exist at the same time
≈
different target

Contribution 3

Spatio-temporal graph



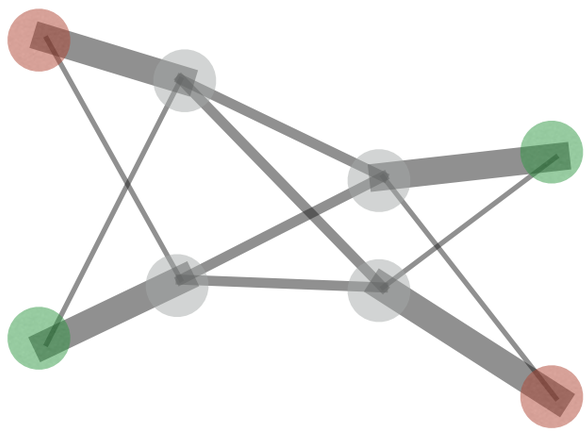
Appearance graph



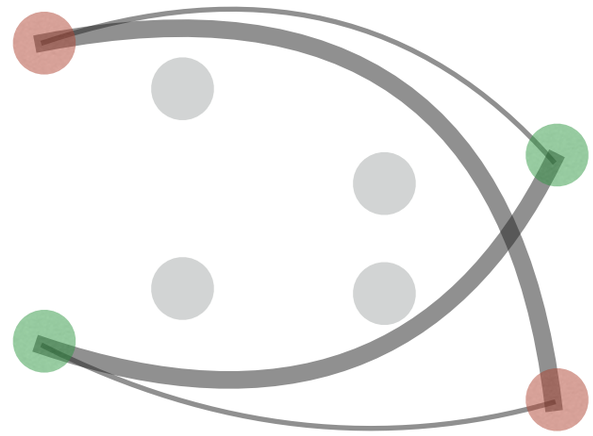
Assign **same** label

Contribution 3

Spatio-temporal graph

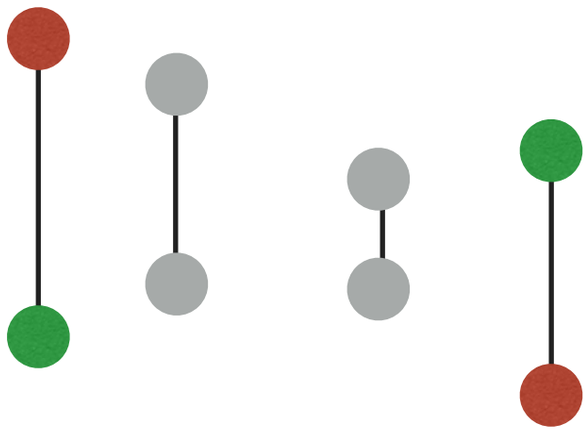


Appearance graph



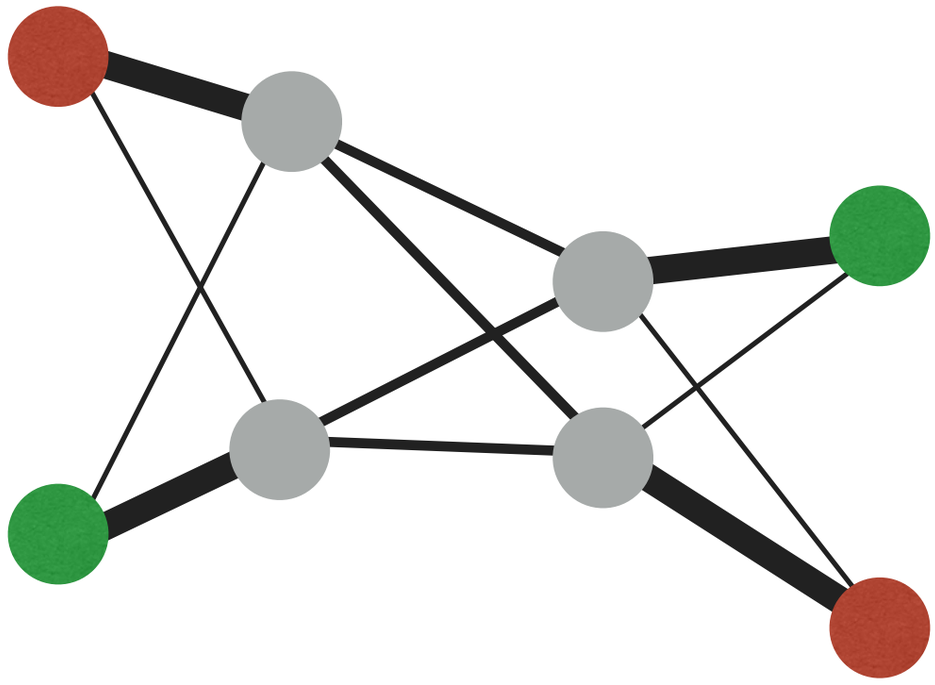
Assign **same** label

Exclusion graph



Assign **different** labels

Contribution 3

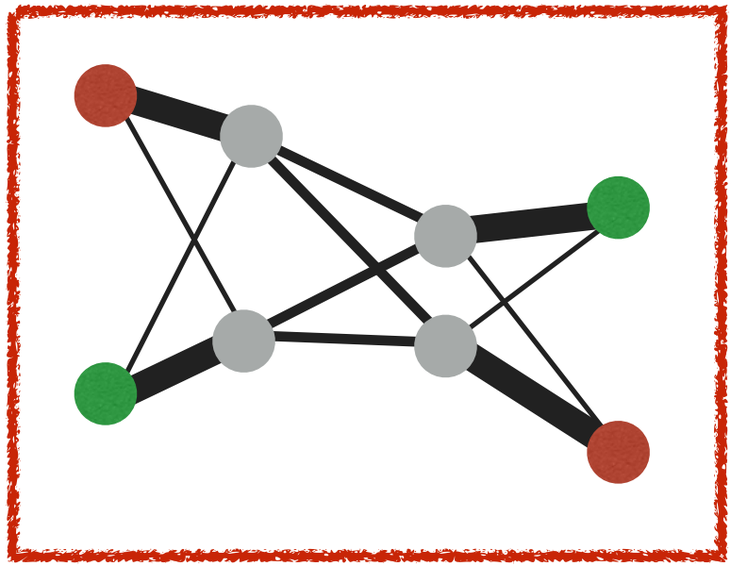


$G = (V, E, W)$ A graph

$Y = (y_1, \cdots, y_n)^\top$ Label assignment

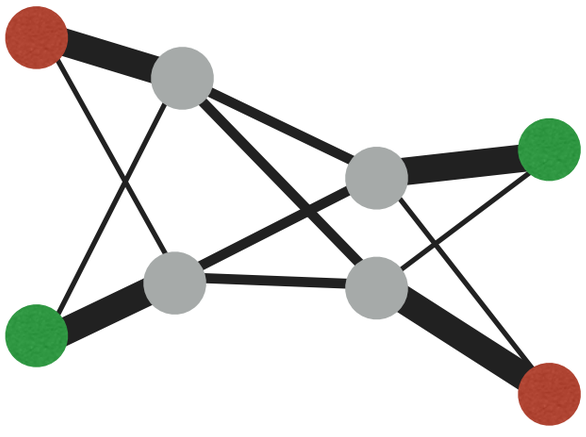
$$E_G(Y) = \frac{1}{2} \sum_{i,j} w_{ij} \|y_i - y_j\|_2^2$$ Labeling energy

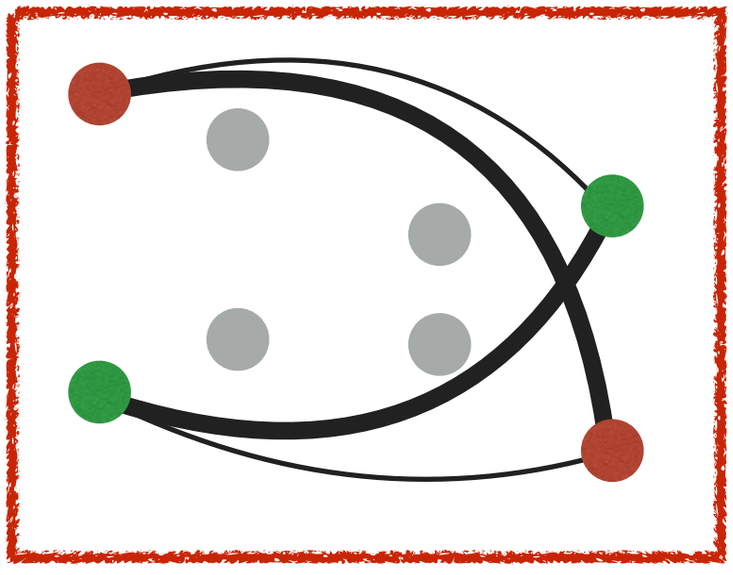
Contribution 3



Labeling energy = E_{spatial}

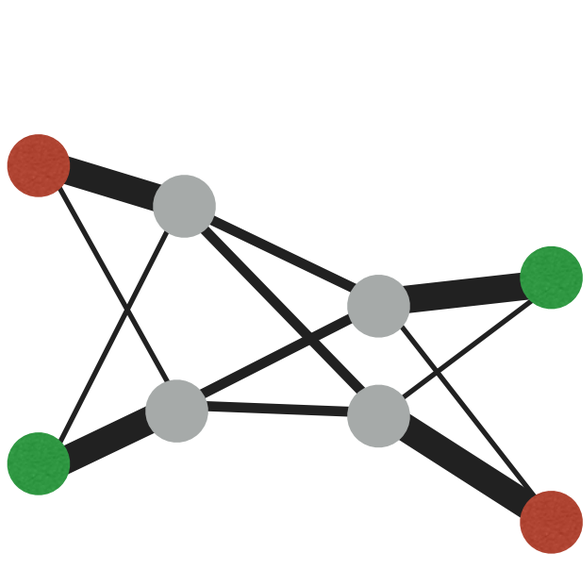
Contribution 3

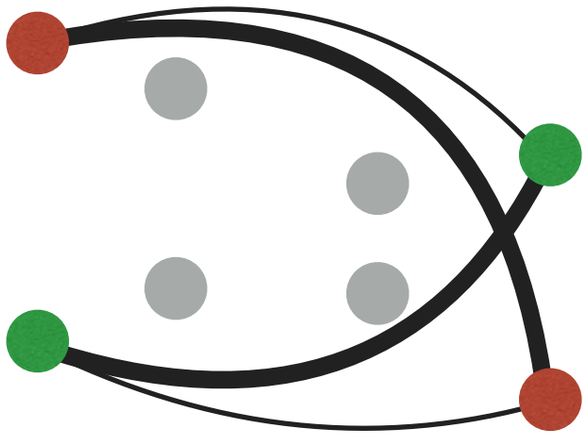


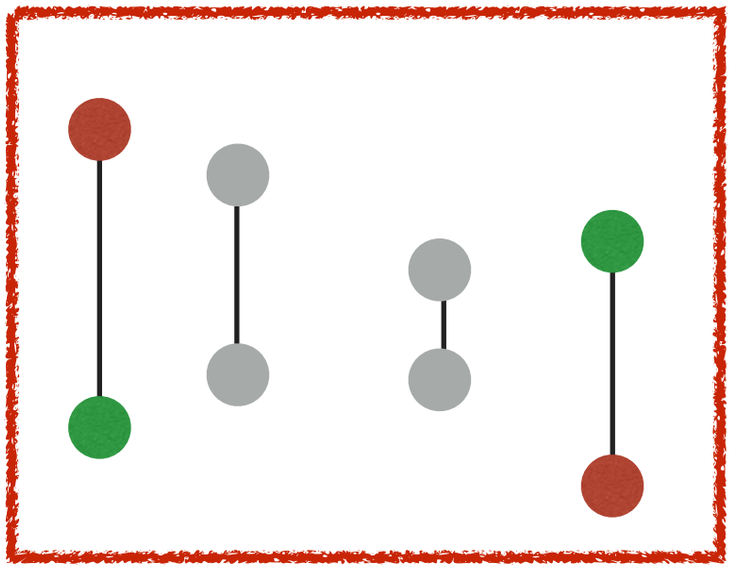


Labeling energy = E_{spatial} + $E_{\text{appearance}}$

Contribution 3







$$\text{Labeling energy} = E_{\text{spatial}} + E_{\text{appearance}} - E_{\text{exclusion}}$$

Difference of convex problem

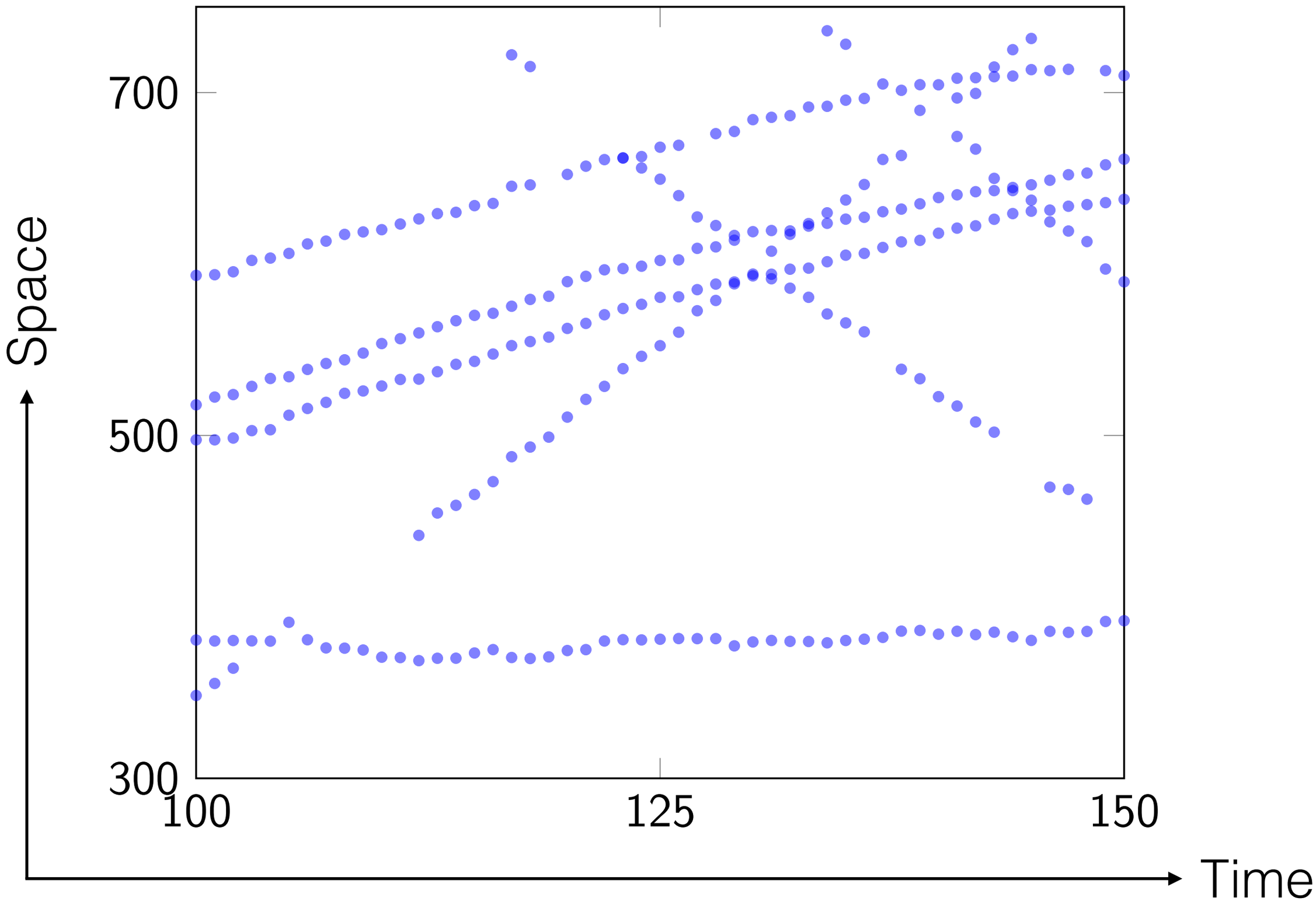
Contribution 3

Solving the problem

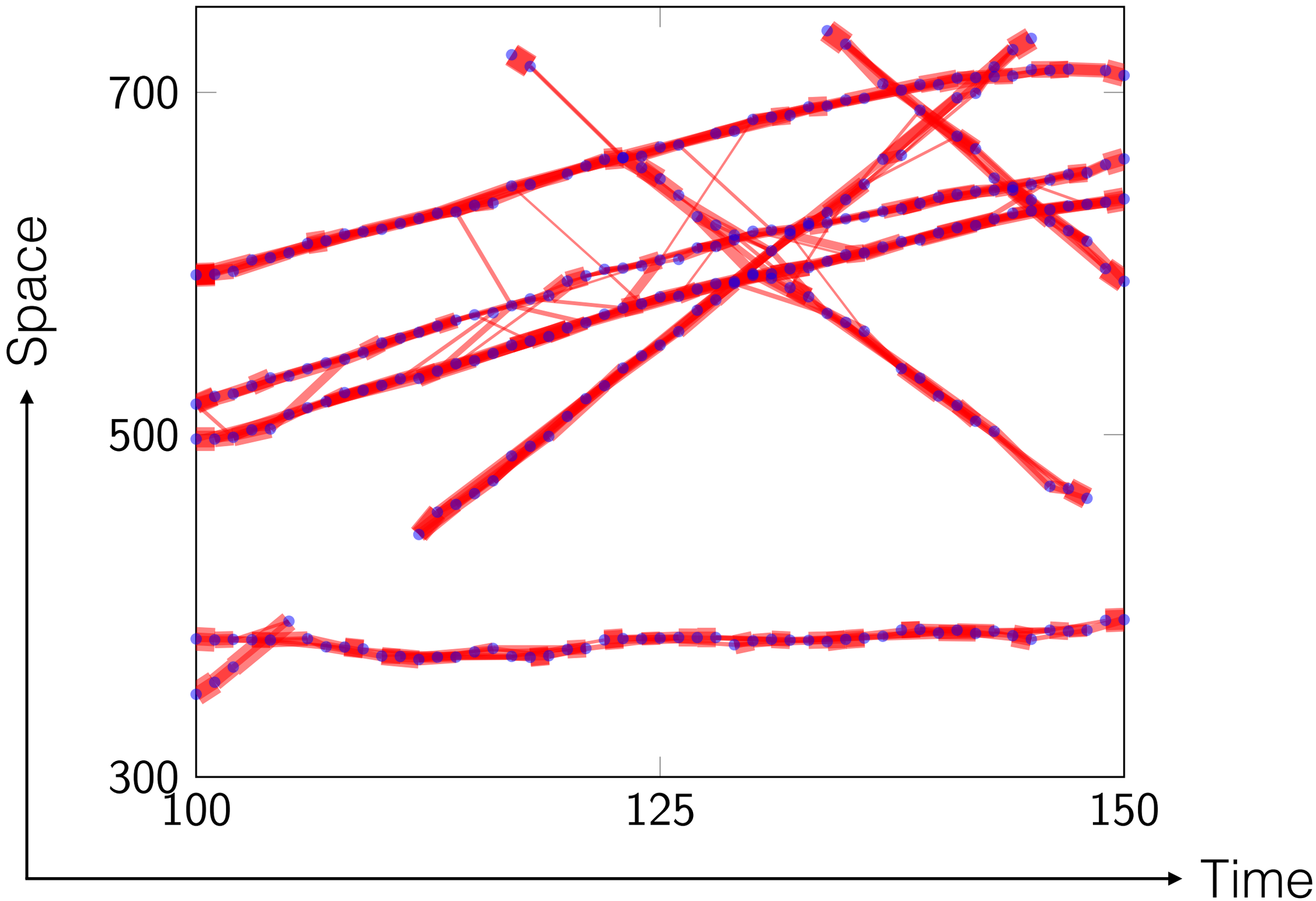
- **Joint**
  - Solve the big problem
- **Node-wise**
  - Solve small problem at each node
  - Possible to handle **online tracking** applications
  - **Reduction in complexity**, possibility of **parallelization**

Demo

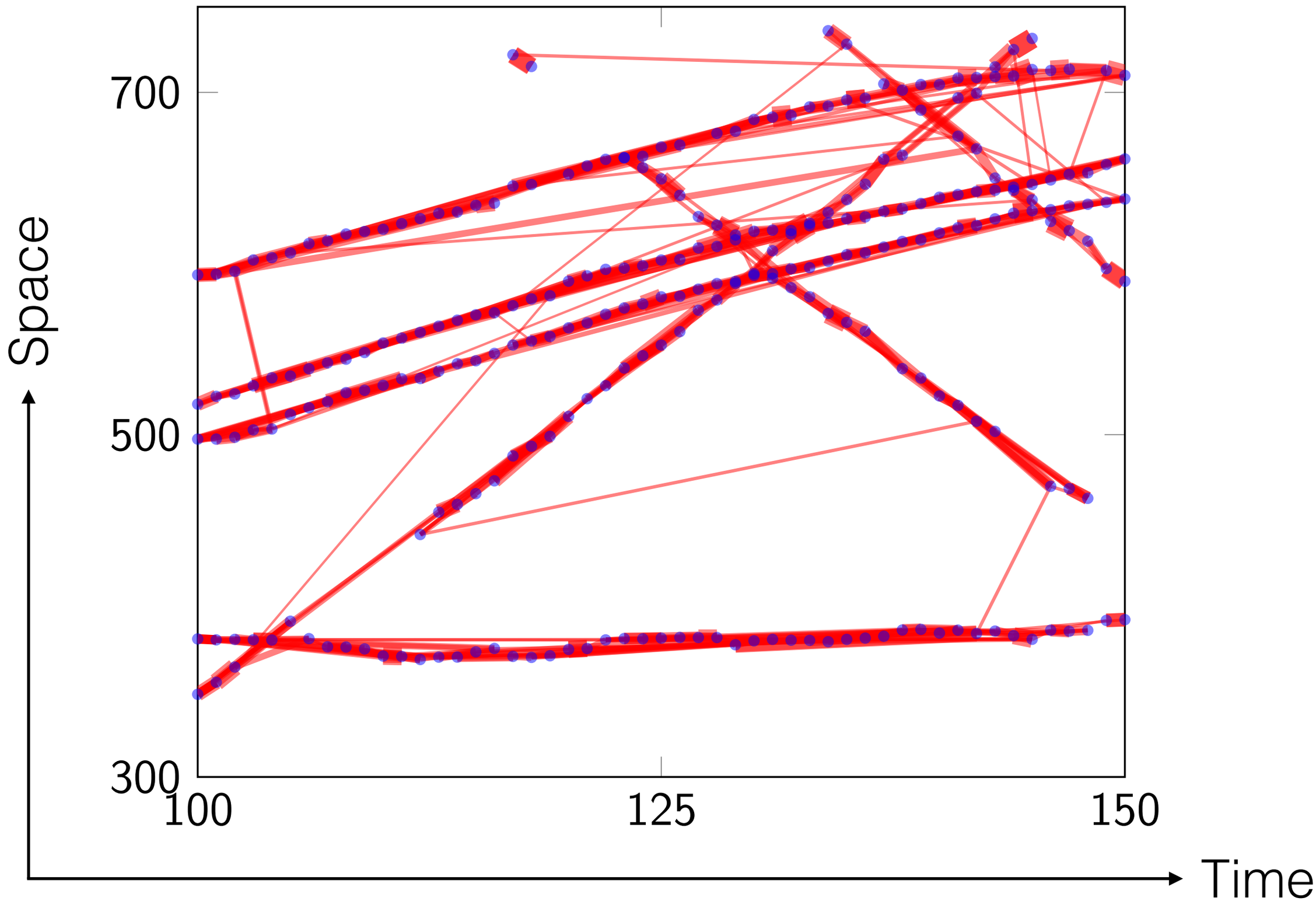
Input detections



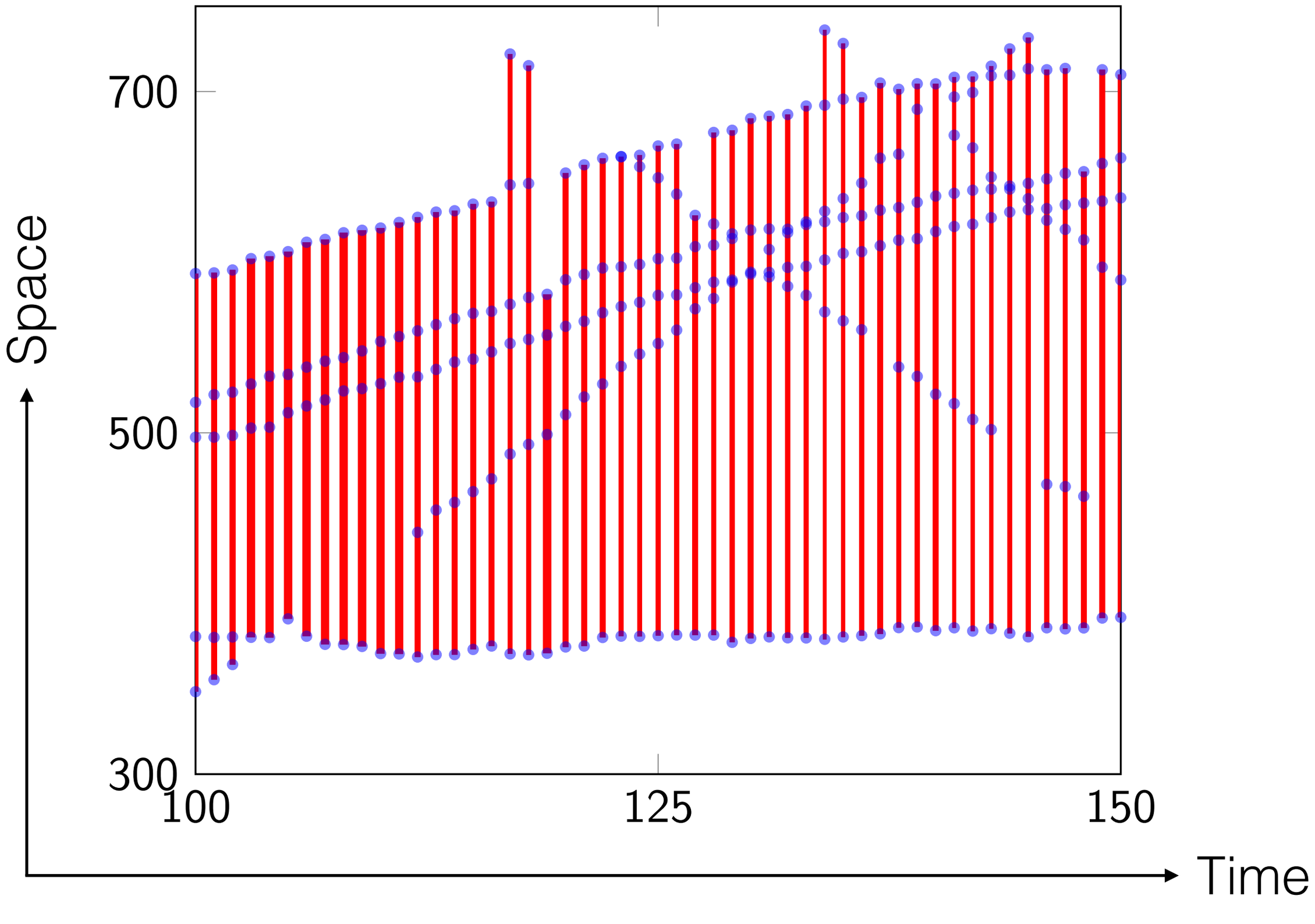
Spatio-temporal graph



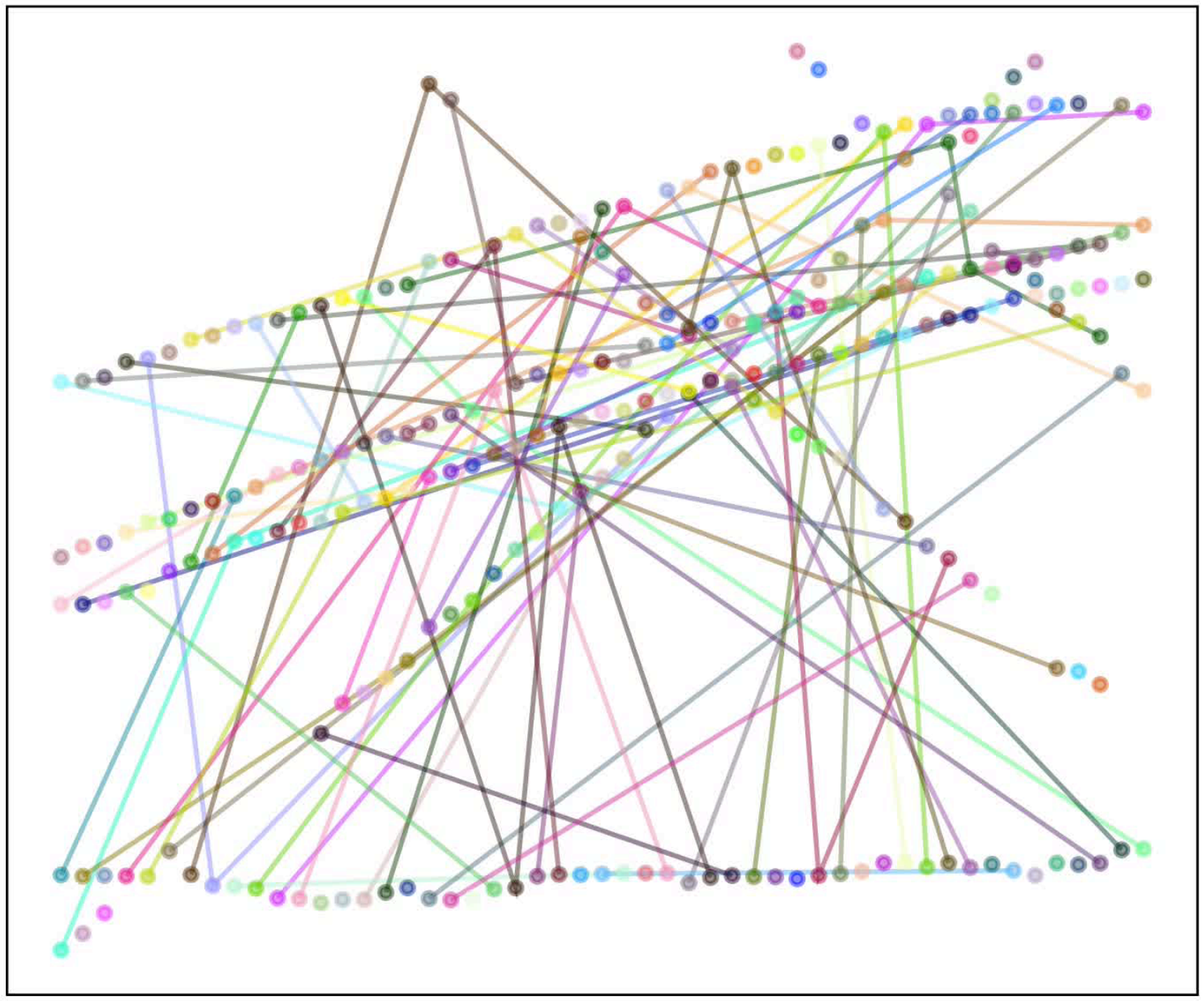
Appearance graph



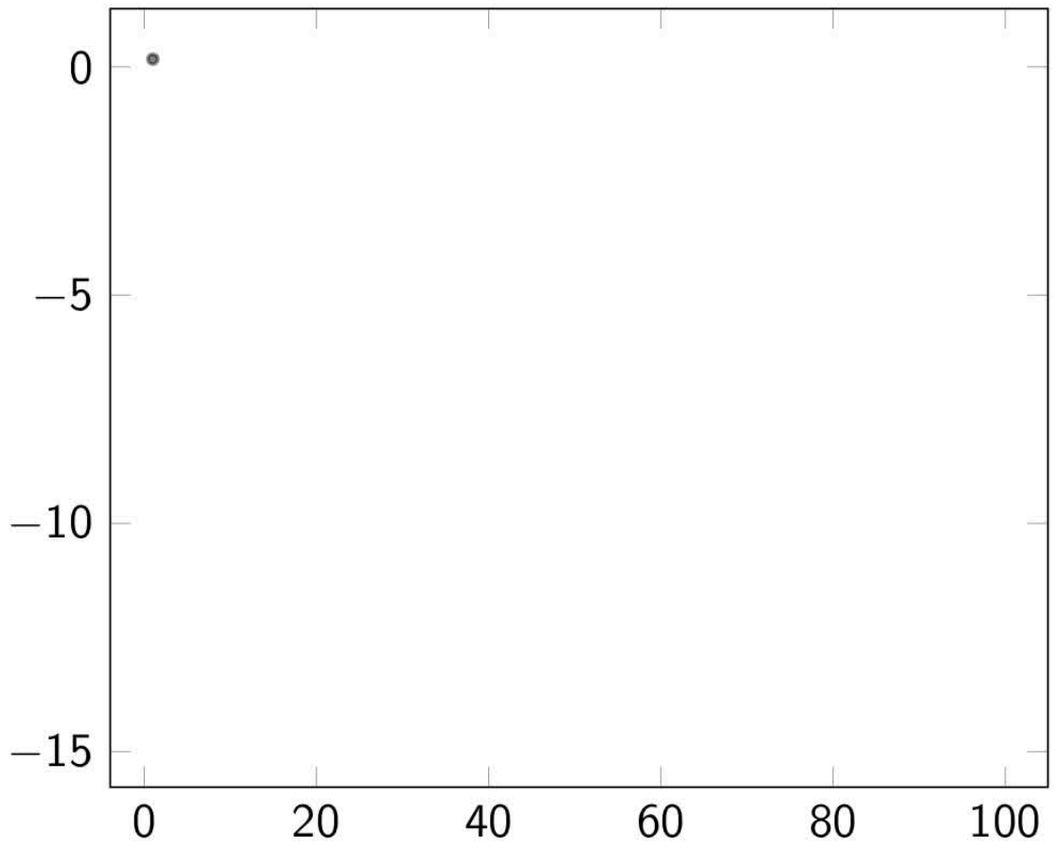
Exclusion graph



Label propagation iterations

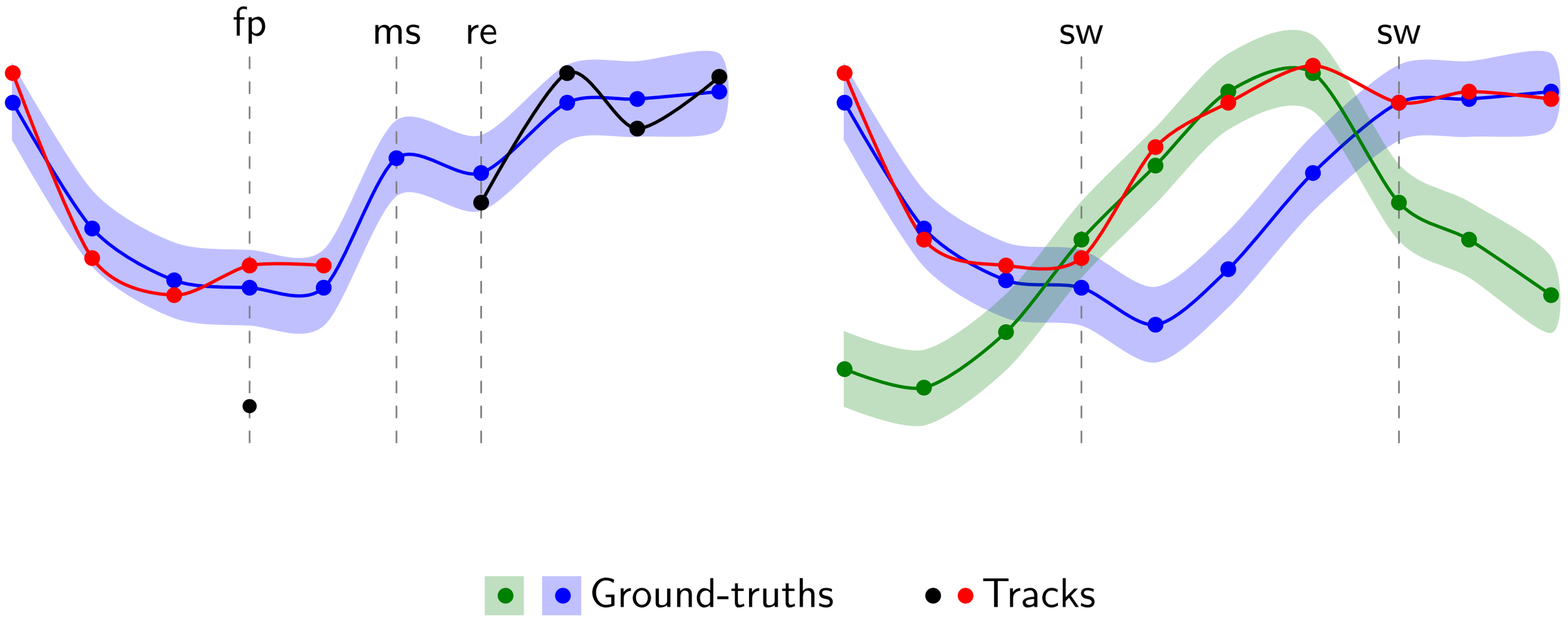


Labeling energy

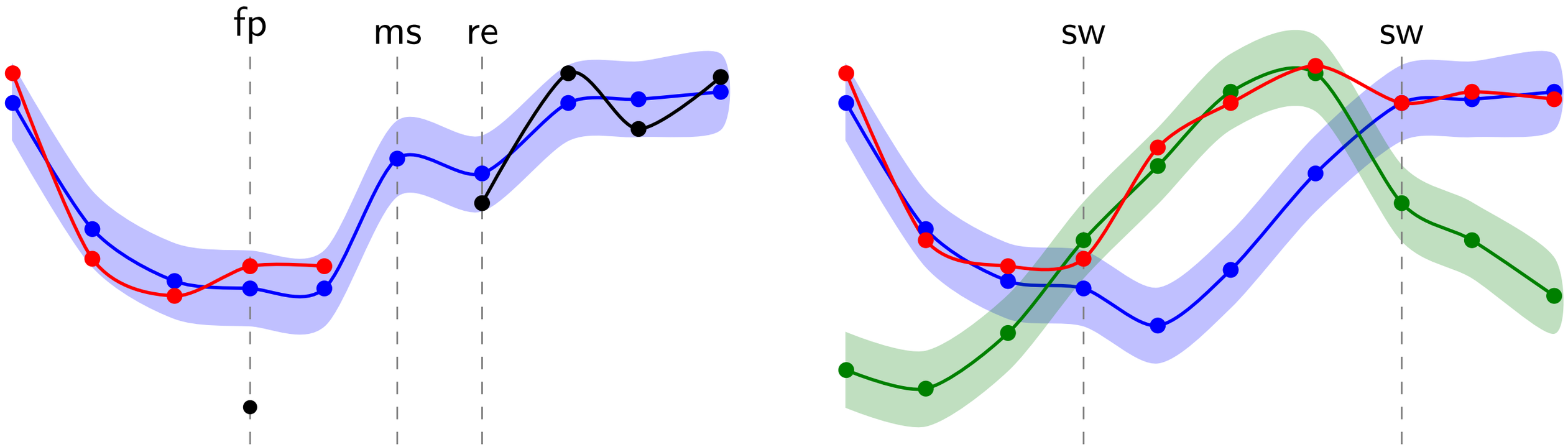


Validations

Multiple Object Tracking Accuracy



Multiple Object Tracking Accuracy



Ground-truths Tracks

$$MOTA = 1 - \frac{FP + MS + RE + SW}{GT}$$

Higher MOTA → Better performance


GARAGE BAUDOUIN PIERRE
Huy - Namur
GRAND HOTEL
de FLANDRE
FEDERALE
PEUGEOT
NAMUR

Method	MOTA (%)	# Identity switches
K-shortest paths	72.91	108
Global appearance constraints	73.07	110
Contribution 1 (IHT)	**86.19**	**12**
Contribution 3 (DLP)	83.80	45

Method	Recognition accuracy (%)
Contribution 2 (PBP)	89.04
Standard belief propagation	73.59

Conclusions

- MOT is **ubiquitous**.

- Previous methods **do not exploit variable reliability** in the appearance features.

Conclusions

- **Iterative hypothesis testing**
  - embeds hypothesis testing to shortest-path computation
- **Prioritized belief propagation**
  - faster convergence
  - better recognition rate
- **Discriminative label propagation**
  - elegant method to exploit various cues
  - requires to solve difference of convex program
  - efficient solution due to node-wise decomposition

Future directions

- **Deployment in different tracking scenarios**
  - evolution of cracks in metals
  - biological cell movement
- **Learn target appearances for long term tracking**
  - exploit the structure of MOT to learn good features to describe the target appearance
- **Learn the graphs for tracking**
  - recent availability of many (diverse) MOT datasets

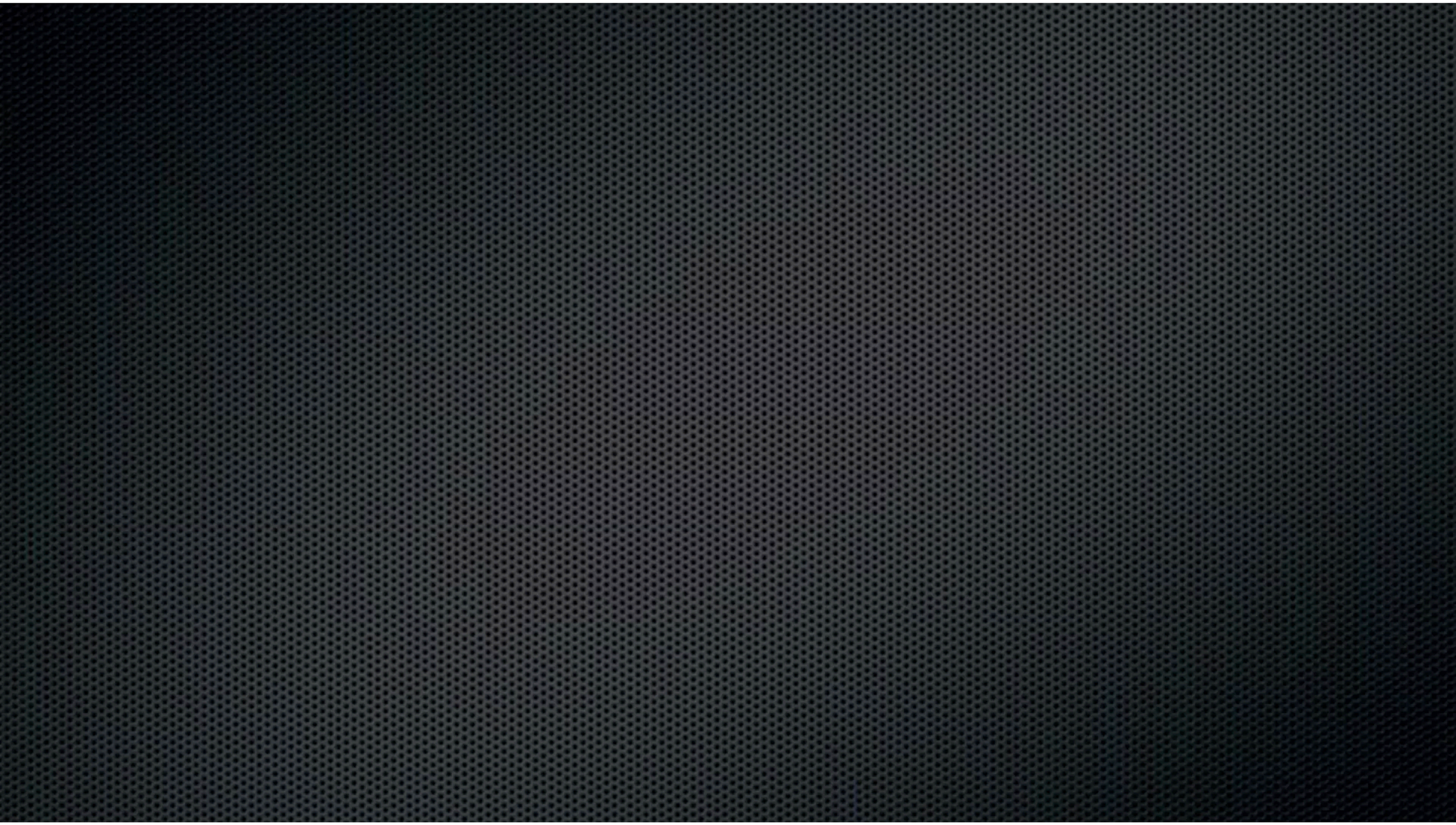
Video sources

- **Cell migration:** http://www.oulu.fi/cse/cmv/bioimage/track

- **Placemeter**: https://vimeo.com/69091237

- **Ant tracking:** https://www.youtube.com/watch?v=YSlEBa4BSb4

Thank you



Acknowledgement

François Fleuret

Jean-François Delaigle

Christophe De Vleeschouwer

Laurent Jacques

Andrea Cavallaro

David Bol

Jean Donia Kaori
Augustin Pascaline Arnaud Cédric
Damien
Stéphanie Adriana
Subir Jeevan Mohieddine
Amir Maxime
Kévin Sébastien François Pierre-Yves
Shambhu Patricia
Valerio Sumit