Unsupervised ML for Classification Problems

John Branch Bedoya
Pedro Atencio Ortiz

December 15, 2020

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Institución
Universitaria
Reacreditada en Alta Calidad

pedroatencio@itm.edu.co

Agenda

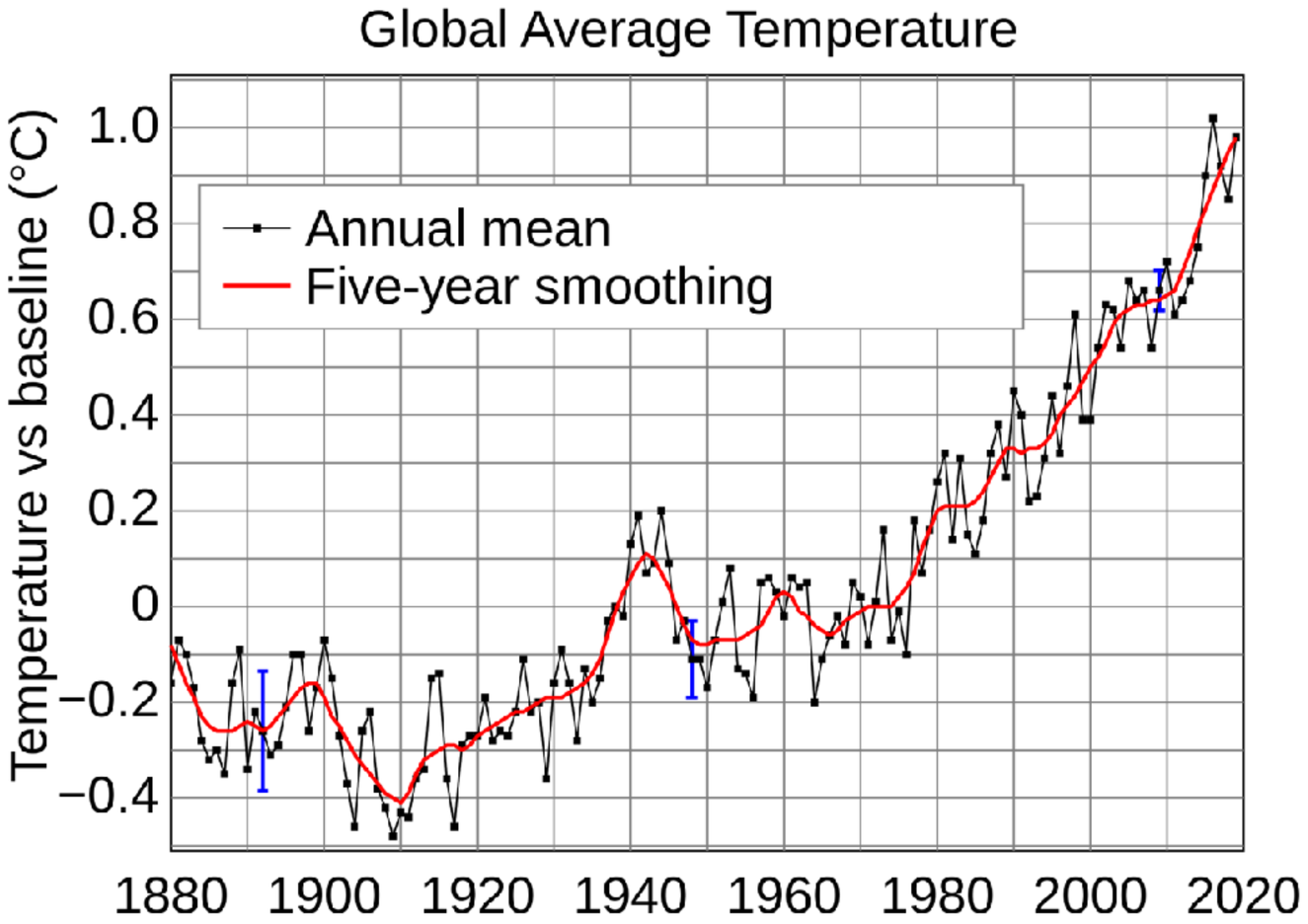
Session 3: Summarization of sequences

1. Sequences: Fundamentals.
2. Sequence Summarization.
    1. Extractive / Abstractive summarization.
    2. Applications.
3. Hierarchical Clustering.
4. Study case.
    1. Video Summarization.
5. Conclusions

1. Sequences: Fundamentals

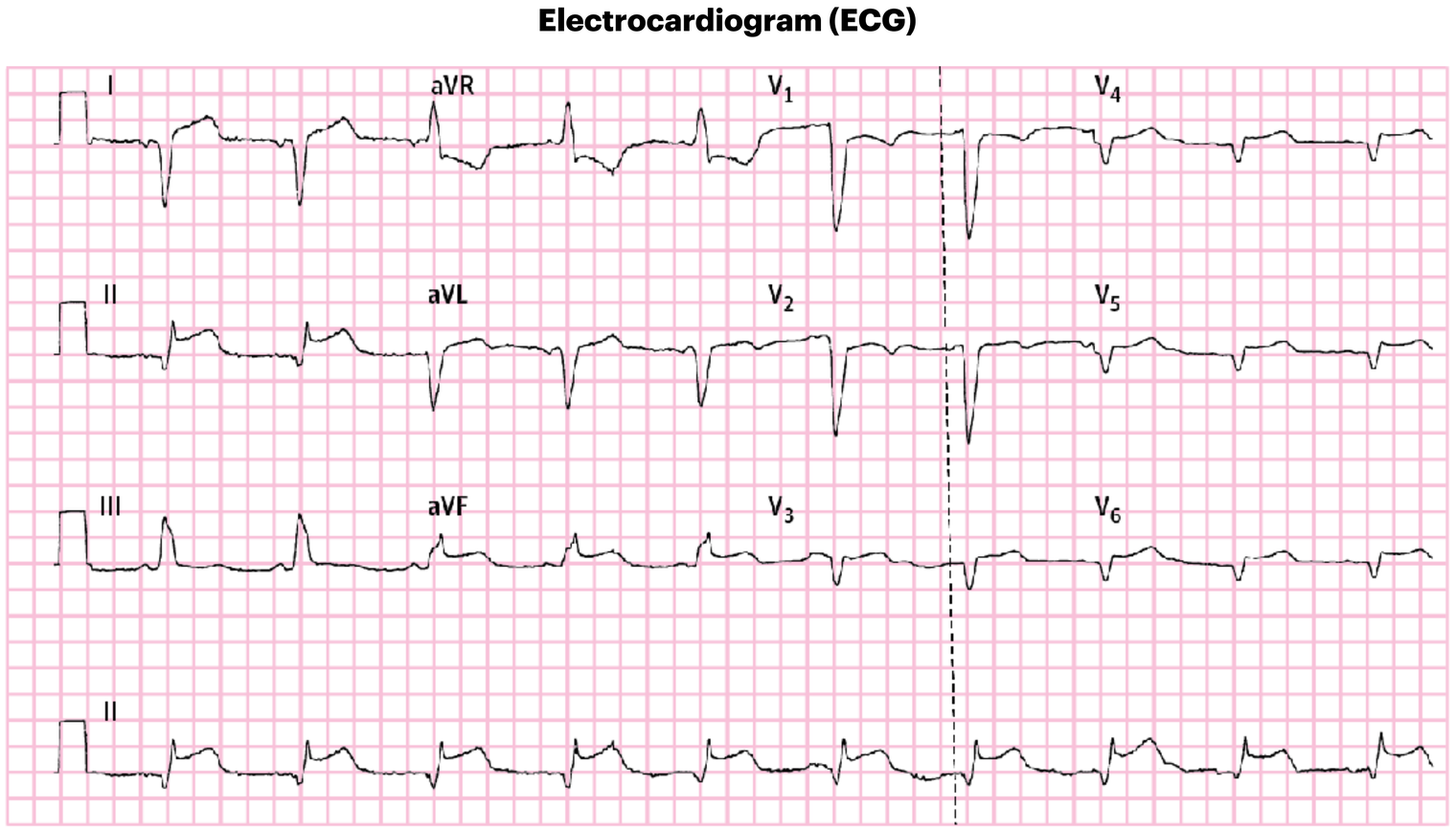
Time series

- Data indexed in time order.
- The concept of sample requires a time window.
- Commonly there is no (X, Y) pair but $(X_{[t-k,t]}, X_{[t+1]})$.
- Time series do not mean regression. Classification in also possible.



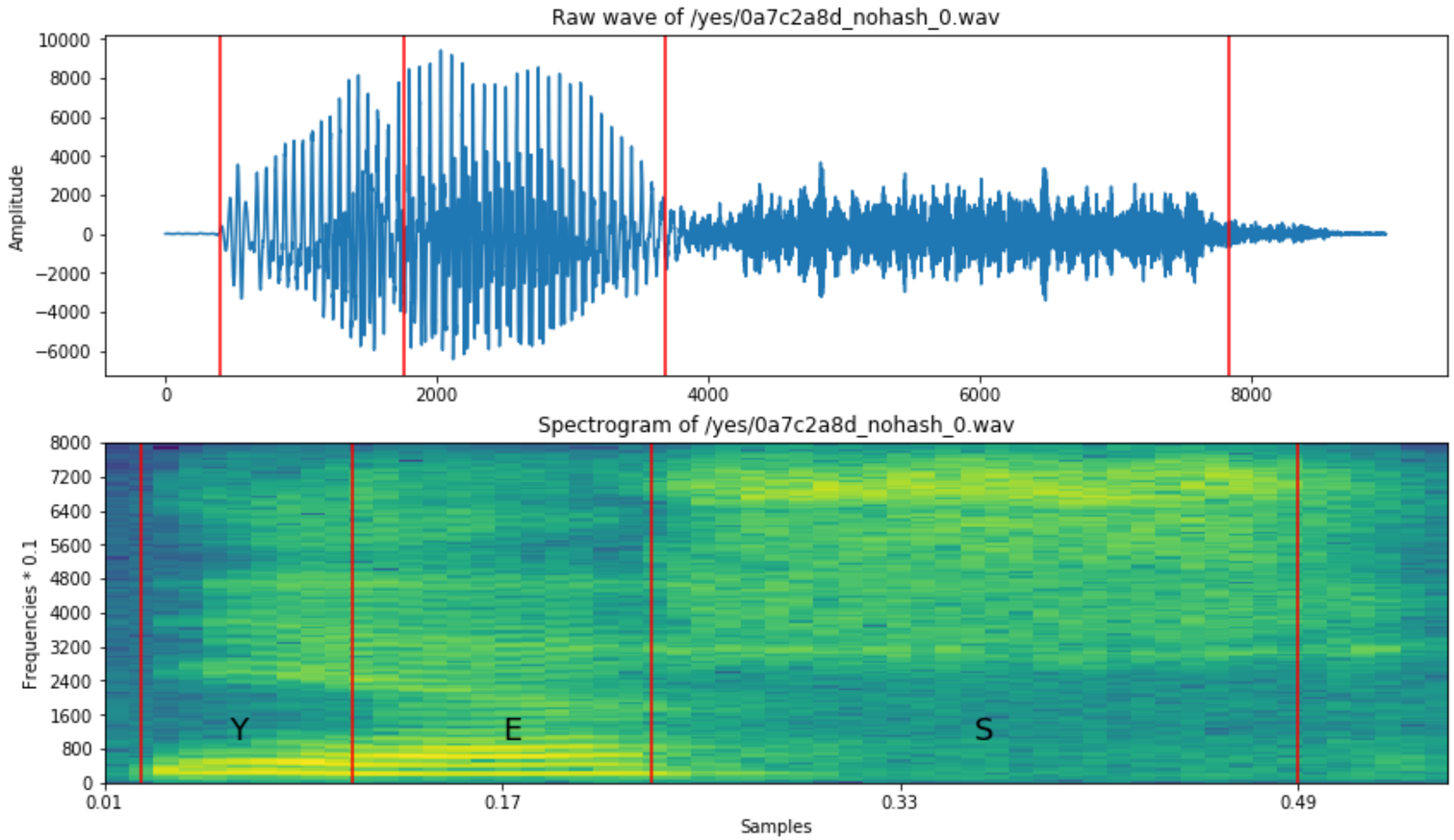
Time series

- Can be univariate or multivariate.
- Can be 1-dimensional o n-dimensional.



Sequence

- Can be univariate or multivariate.
- Can be 1-dimensional o n-dimensional.

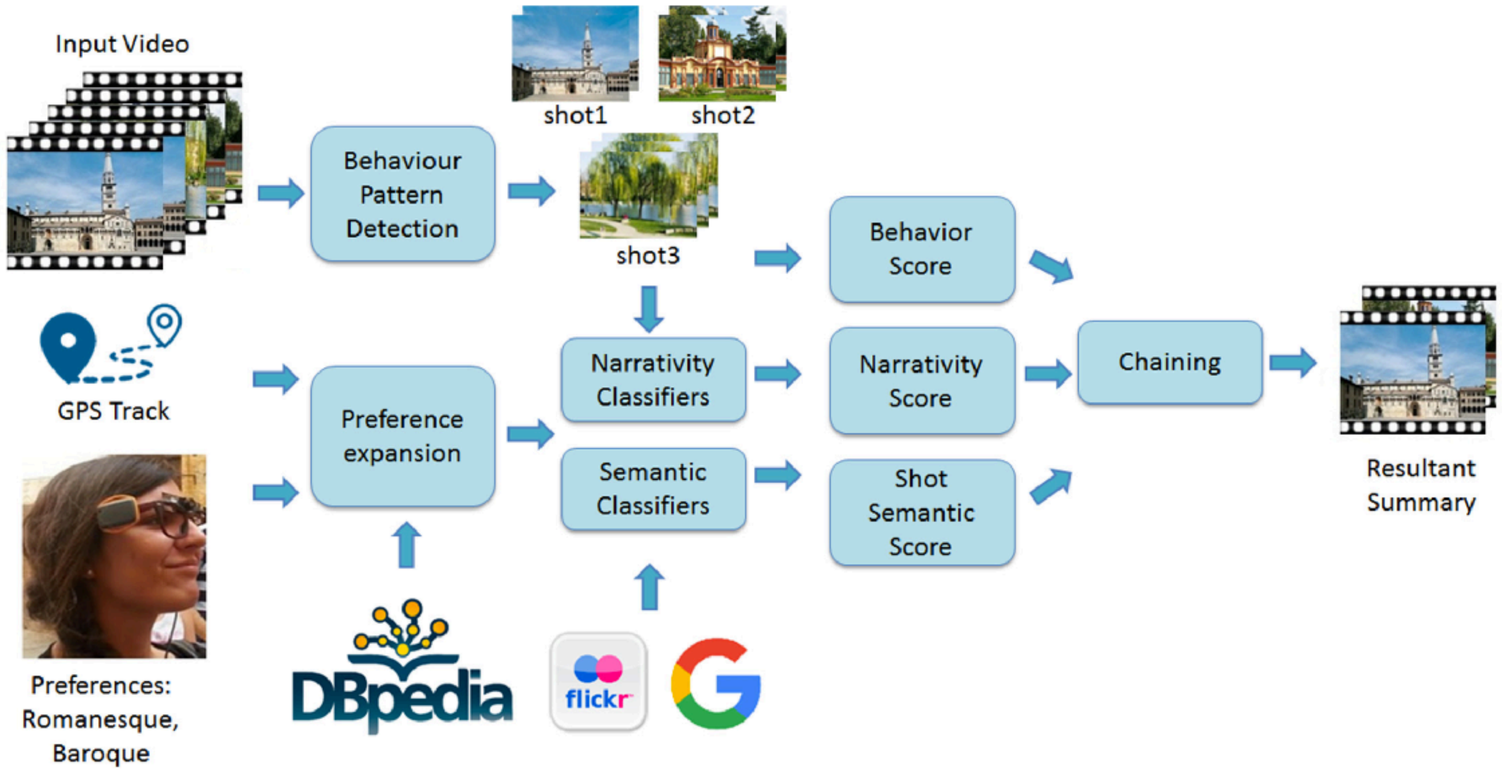


Sequence

- Time is not always clear, but order.
- Words in a text can be considered a sequence, not in time but in order.

Sequence

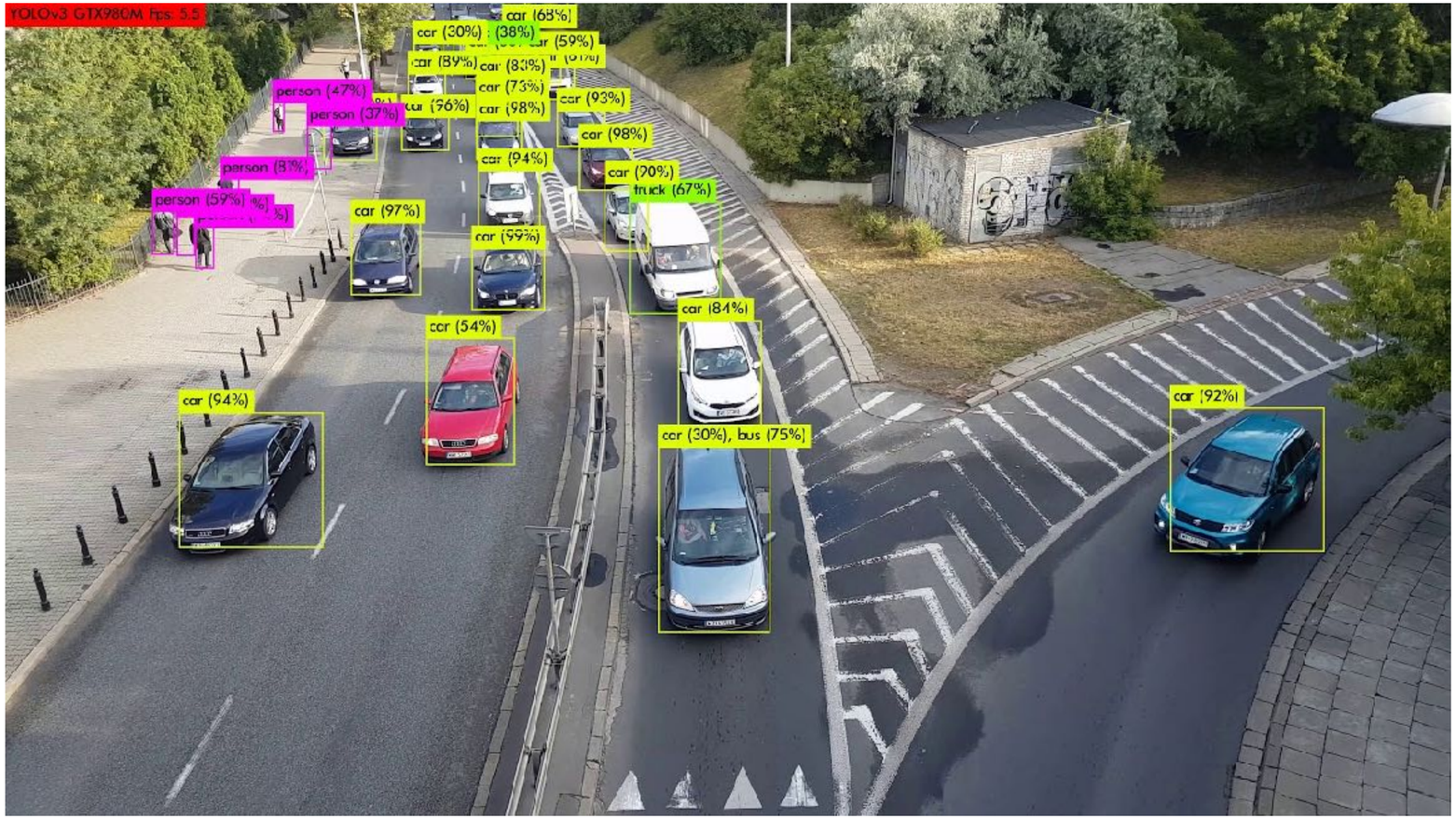
- Can be multimodal.
- **Multimodality**: same phenomena, observations from different sensors / sources.



(Varini et.al., 2017) Personalized Egocentric Video Summarization of Cultural Tour on User Preferences Input

Sequence

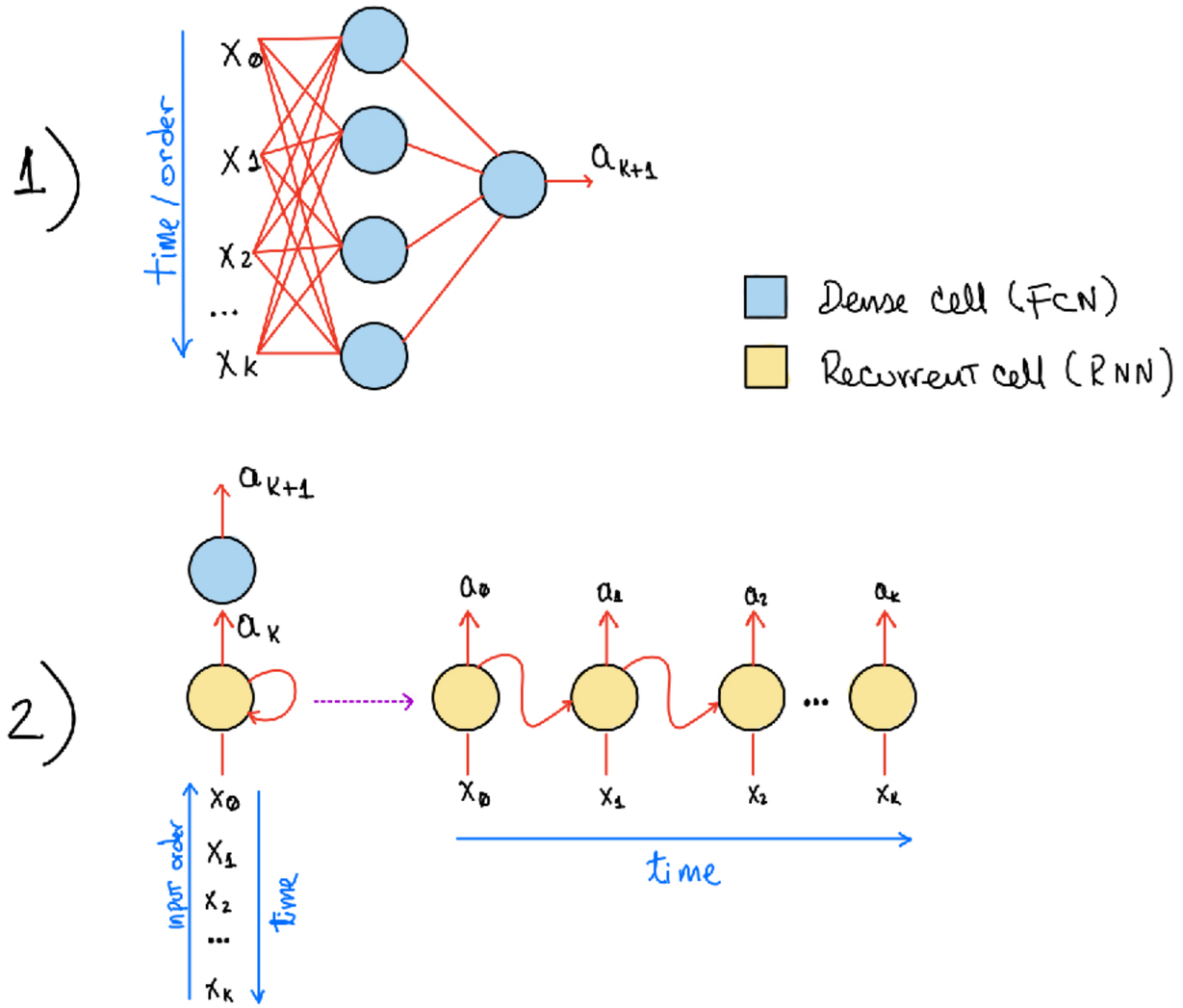
- Sometimes simplification, compression or summarization is required.
- e.g. Bounding boxes and labels instead of images in order to predict accidents.



Sequence processing

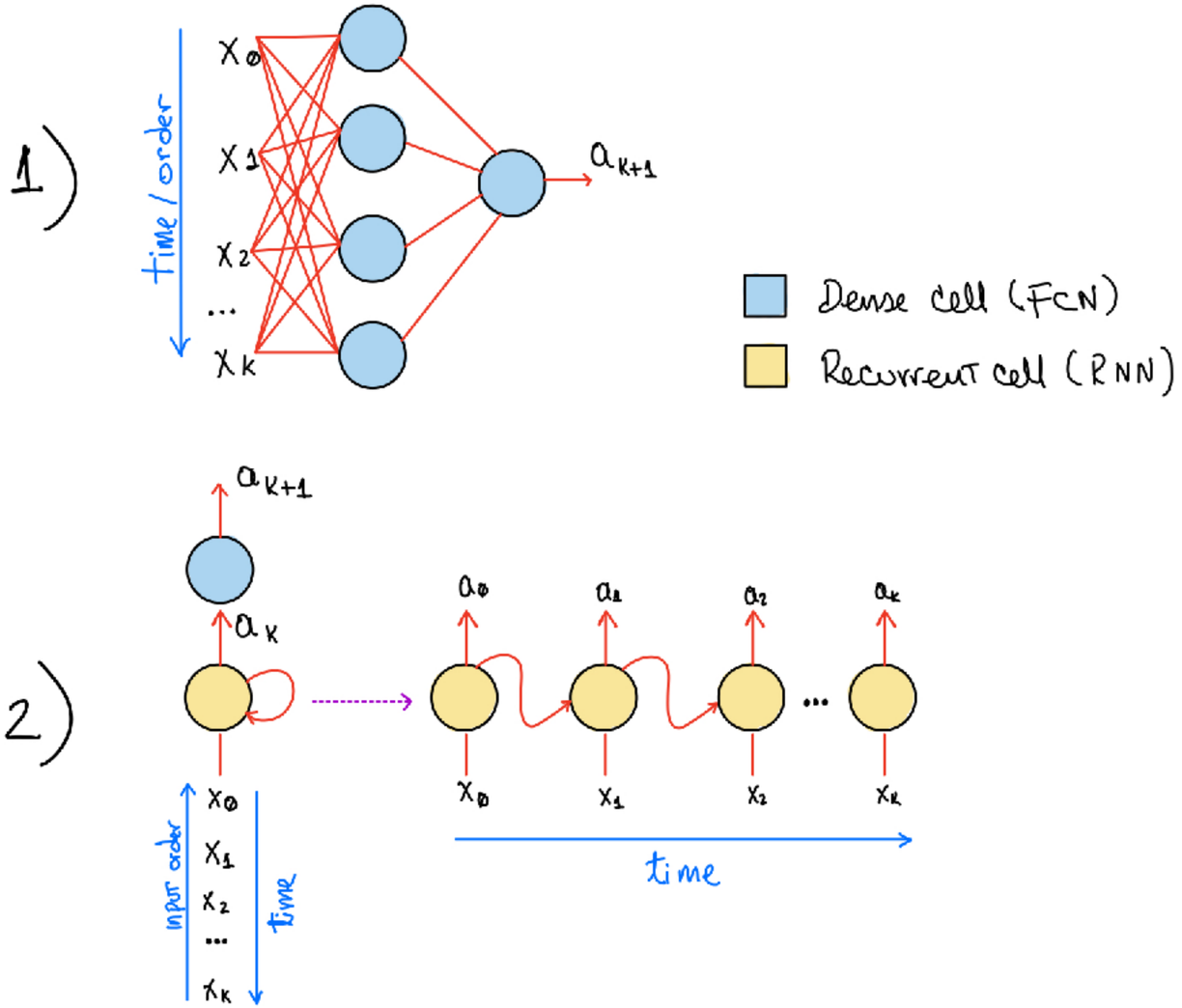
- 1) Each sample x_i in a serie is taken as an independent input of a FCN (dense) network.

- 2) Each sample x_i in a serie is sequentially inputted in a recurrent layer. The output of the recurrent layer is taken as the input of a dense layer.



Sequence processing

[Explore time series processing with Keras.]



Study case: Video classification

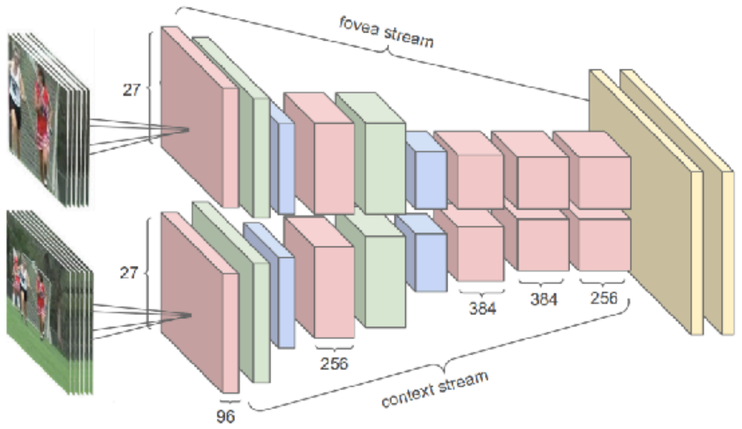
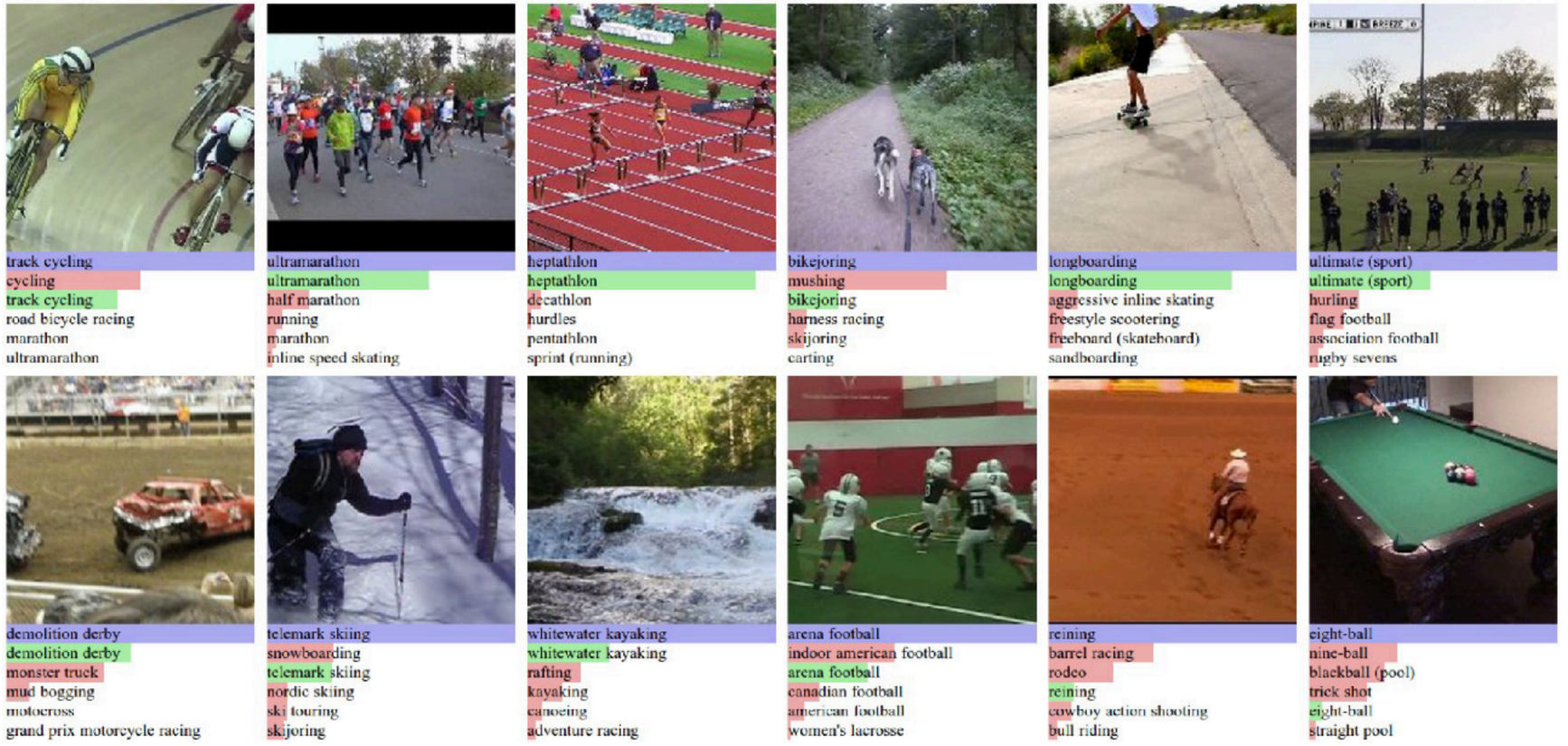


Figure 2: Multiresolution CNN architecture. Input frames are fed into two separate streams of processing: a *context stream* that models low-resolution image and a *fovea stream* that processes high-resolution center crop. Both streams consist of alternating convolution (red), normalization (green) and pooling (blue) layers. Both streams converge to two fully connected layers (yellow).

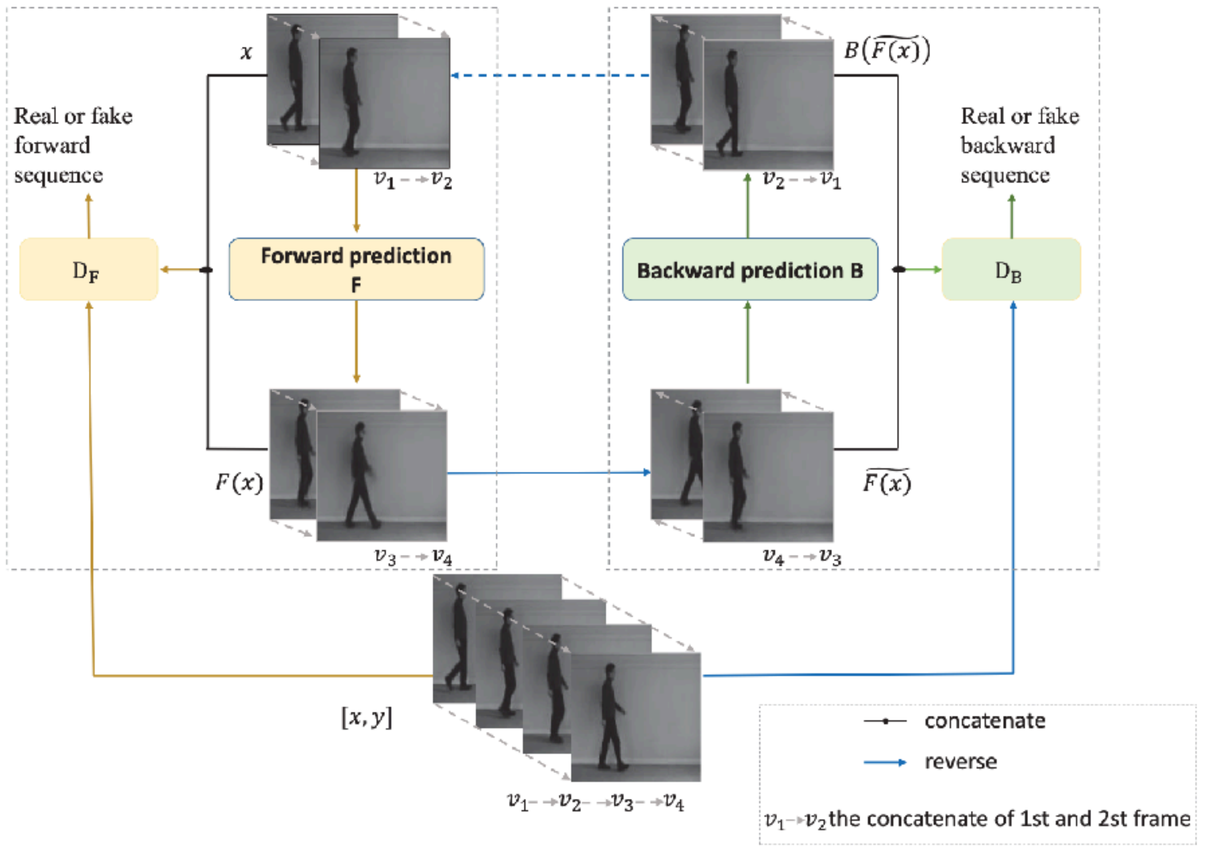
(Karpathy et. al. , 2014) Large-scale Video Classification with Convolutional Neural Networks.

Study case: Video classification



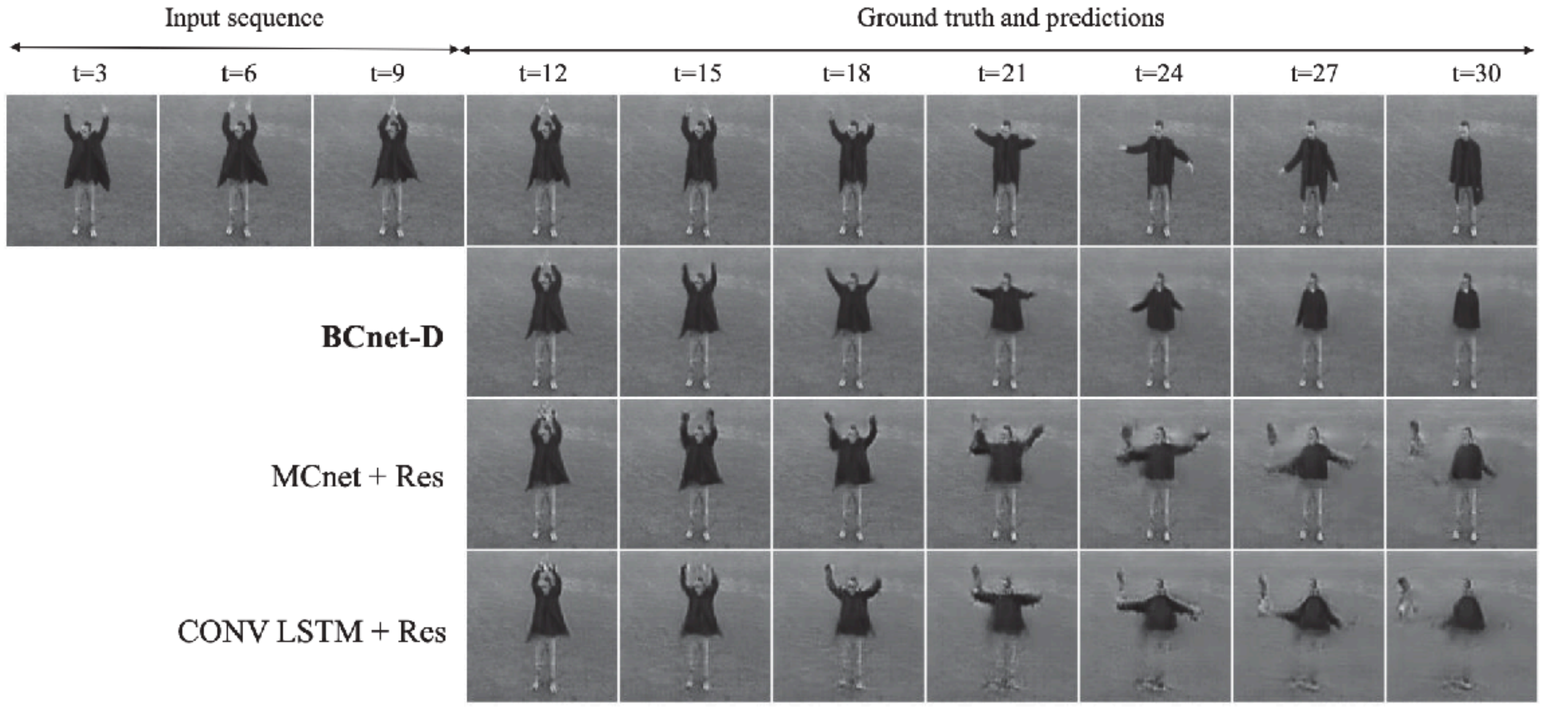
(Karpathy et. al. , 2014) Large-scale Video Classification with Convolutional Neural Networks.

Study case: Video prediction



(Hou et. al. , 2019) Video Prediction with Bidirectional Constraint Network .

Study case: Video prediction



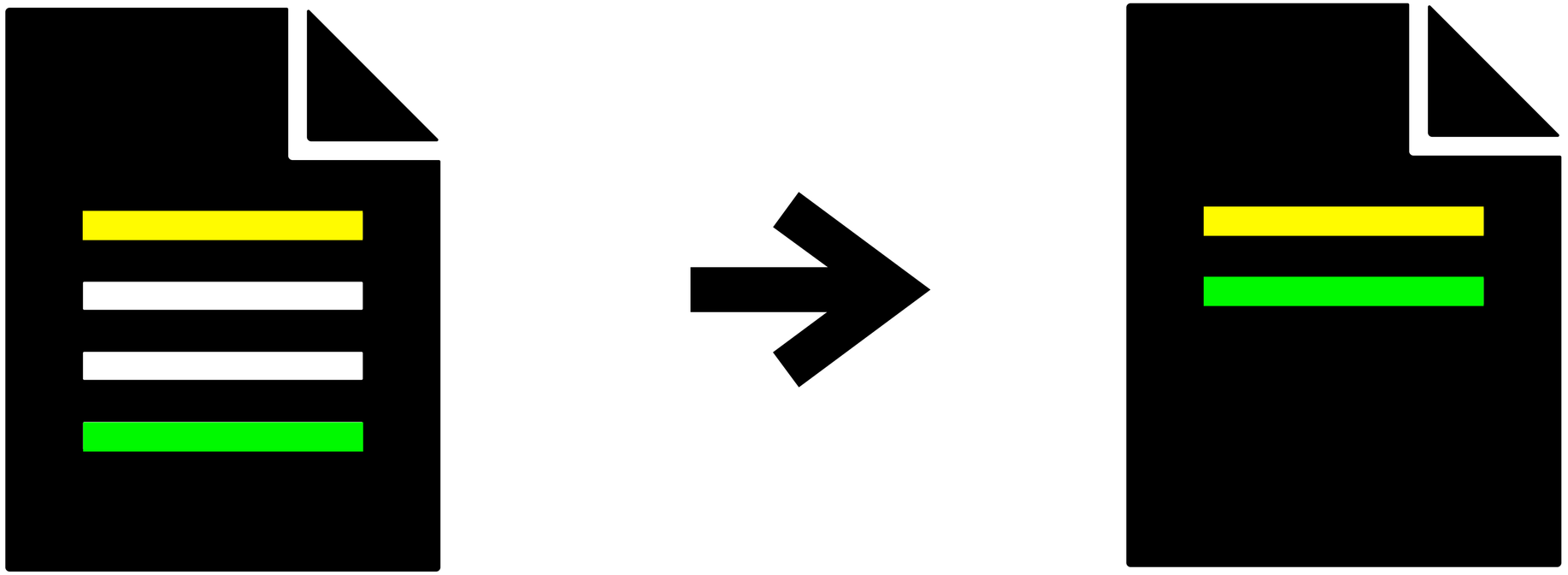
(b) Hand-Waving

Fig. 8. Prediction samples from KTH dataset. We display predictions starting from the 12^{th} frame, in every 3 time steps.

(Hou et. al. , 2019) Video Prediction with Bidirectional Constraint Network .

2. Sequence Summarization

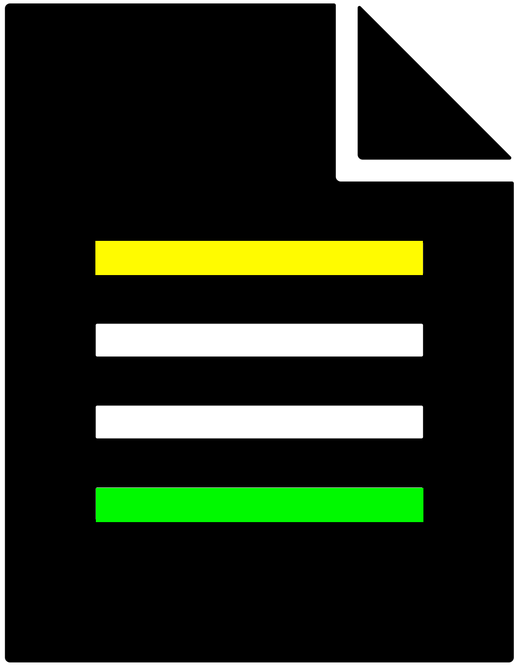
Extractive Summarization



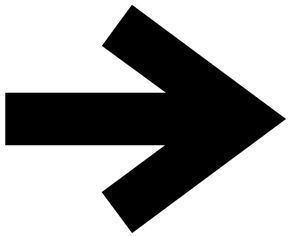
Input text

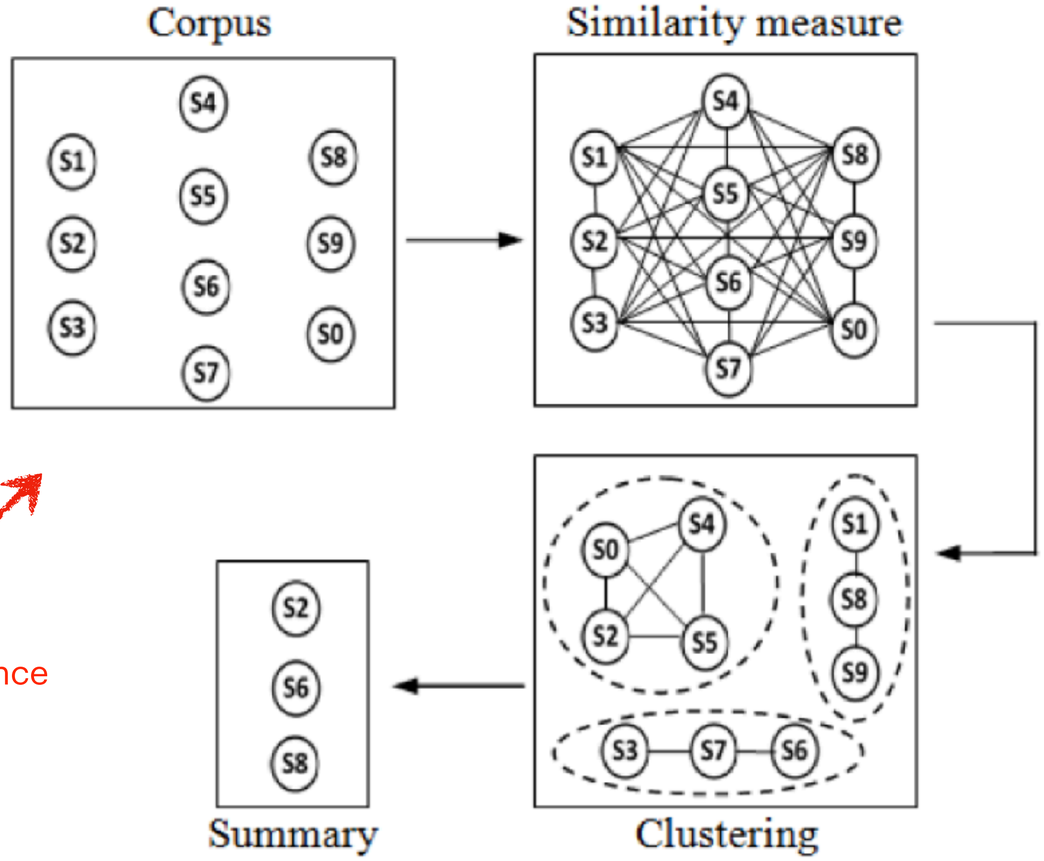
Summary

Extractive Summarization



Input text

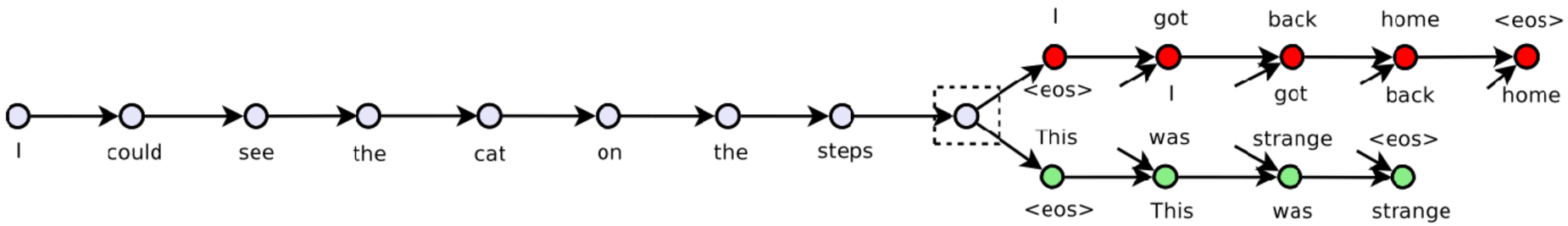




It is required a sentence representation

(Kumar et. al. , 2016) A Review on Automatic Text Summarization Approaches.

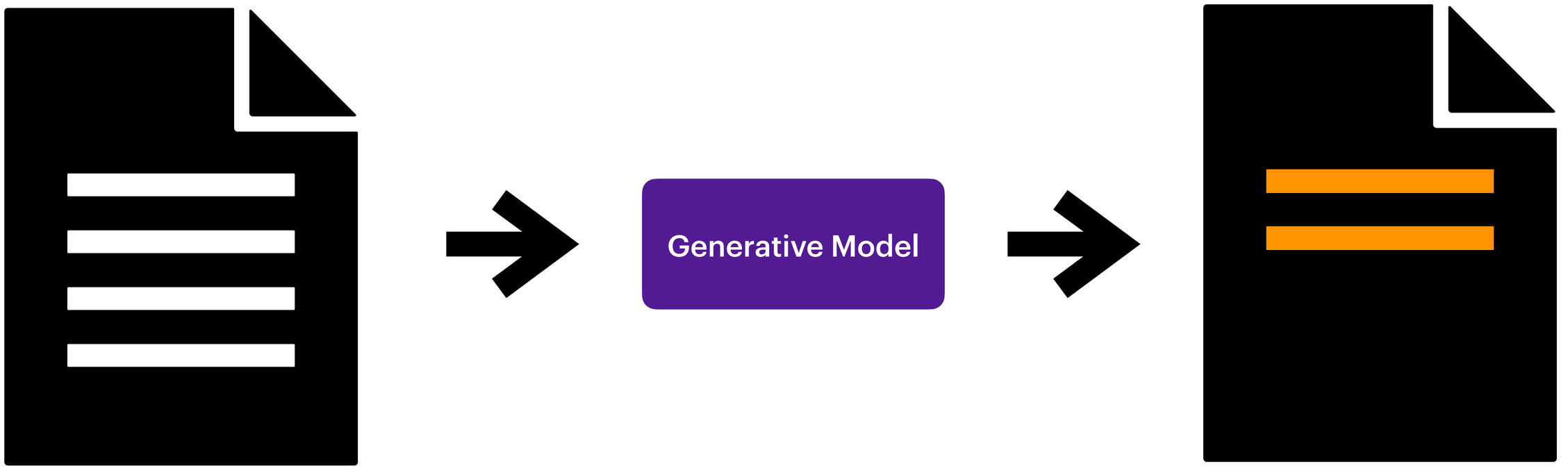
Sentence Embedding: Skip-thought vectors



Query and nearest sentence
he ran his hand inside his coat , double-checking that the unopened letter was still there . he slipped his hand between his coat and his shirt , where the folded copies lay in a brown envelope .
im sure youll have a glamorous evening , she said , giving an exaggerated wink . im really glad you came to the party tonight , he said , turning to her .
although she could tell he had n't been too invested in any of their other chitchat , he seemed genuinely curious about this . although he had n't been following her career with a microscope , he 'd definitely taken notice of her appearances .
an annoying buzz started to ring in my ears , becoming louder and louder as my vision began to swim . a weighty pressure landed on my lungs and my vision blurred at the edges , threatening my consciousness altogether .

(Kiros et. al. , 2015) Skip-thought vectors.

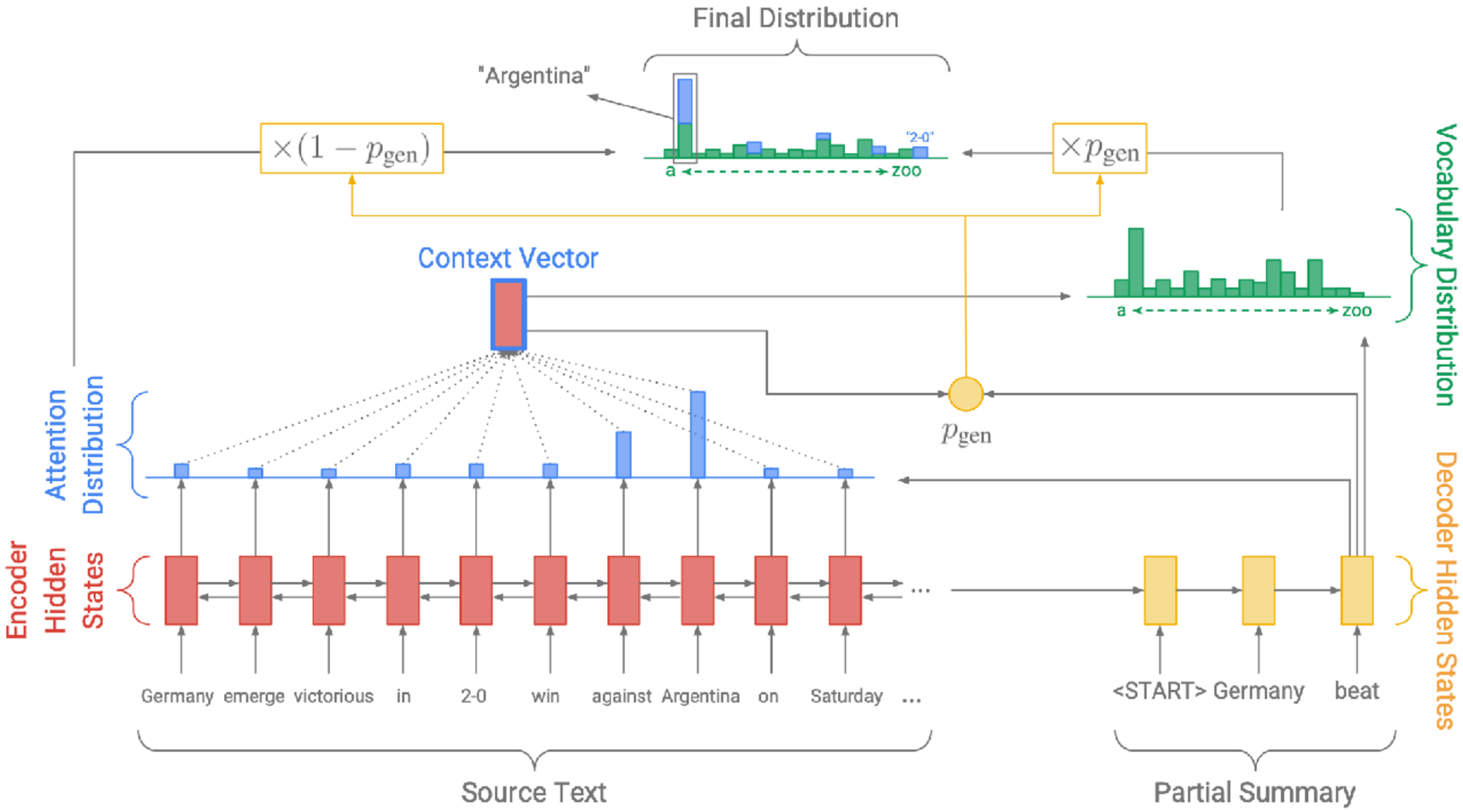
Abstractive Summarization



Input text

Summary

Abstractive Summarization



(See et. al. , 2017) Get To The Point: Summarization with Pointer-Generator Networks.

Abstractive Summarization

Original Text (truncated): lagos, nigeria (cnn) a day after winning nigeria's presidency, ***muhammadu buhari*** told cnn's christiane amanpour that **he plans to aggressively fight corruption that has long plagued nigeria** and go after the root of the nation's unrest. *buhari* said he'll "rapidly give attention" to curbing violence in the northeast part of nigeria, where the terrorist group boko haram operates. by cooperating with neighboring nations chad, cameroon and niger, **he said his administration is confident it will be able to thwart criminals** and others contributing to nigeria's instability. for the first time in nigeria's history, the opposition defeated the ruling party in democratic elections. *buhari* defeated incumbent goodluck jonathan by about 2 million votes, according to nigeria's independent national electoral commission. **the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.**

Baseline Seq2Seq + Attention: **UNK UNK** says his administration is confident it will be able to **destabilize nigeria's economy**. **UNK** says his administration is confident it will be able to thwart criminals and other **nigerians**. **he says the country has long nigeria and nigeria's economy.**

Pointer-Gen: *muhammadu buhari* says he plans to aggressively fight corruption **in the northeast part of nigeria**. he says he'll "rapidly give attention" to curbing violence **in the northeast part of nigeria**. he says his administration is confident it will be able to thwart criminals.

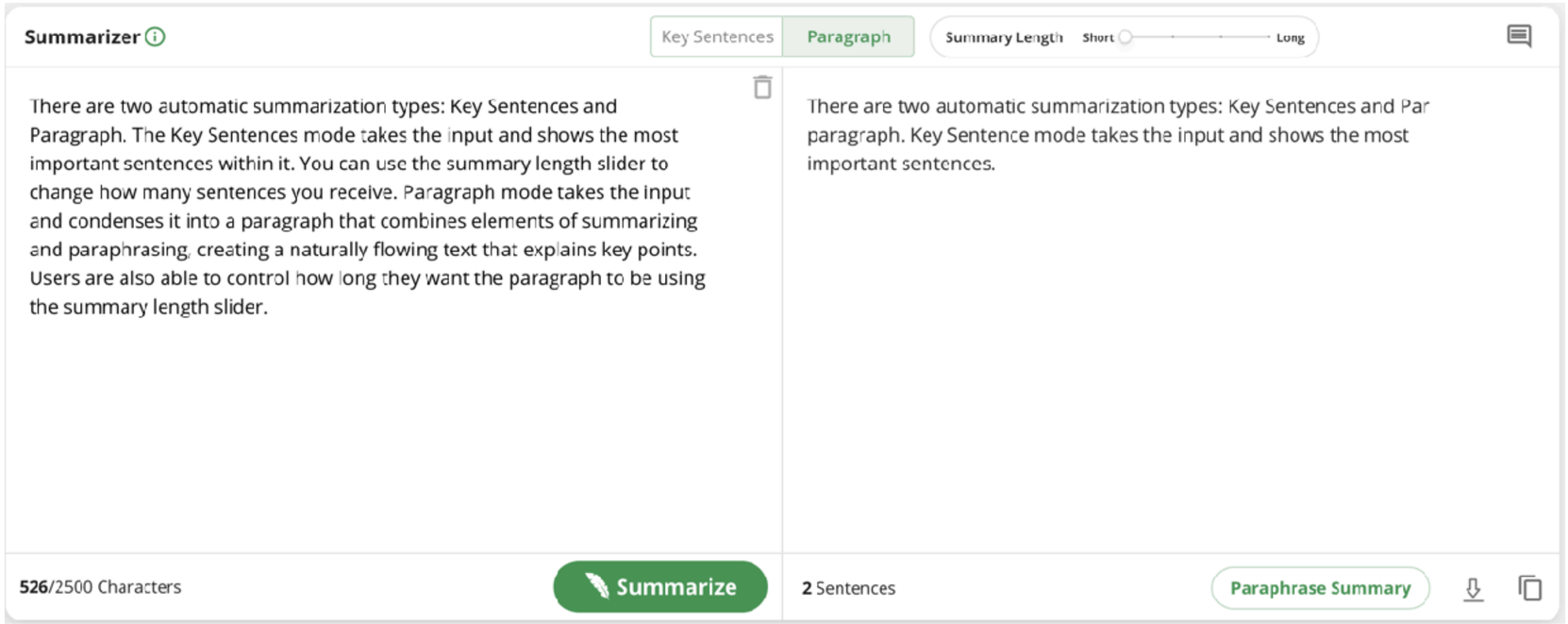
Pointer-Gen + Coverage: *muhammadu buhari* says he plans to aggressively fight corruption that has long plagued nigeria. he says his administration is confident it will be able to thwart criminals. the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.

(See et. al. , 2017) Get To The Point: Summarization with Pointer-Generator Networks.

Applications

Text summarization

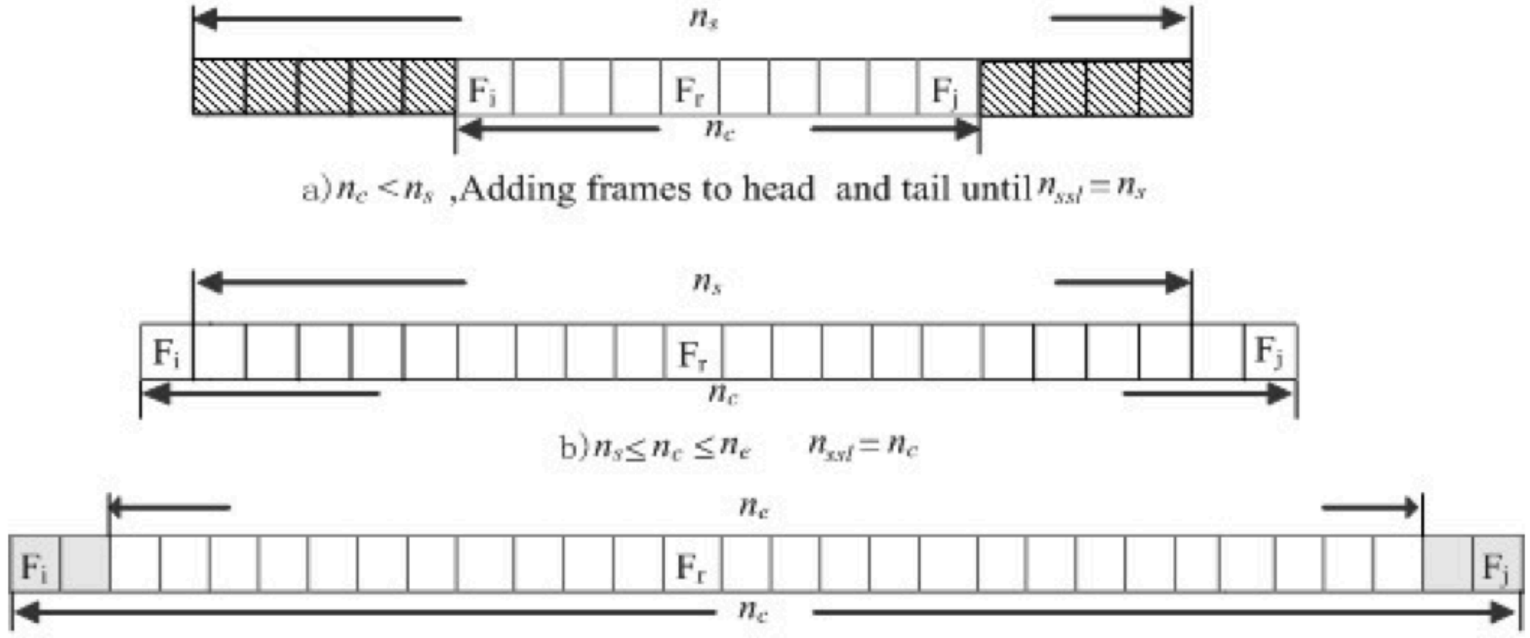
- **Input**: arbitrary-length text.
- **Output**: shorter text.
- **Applications**: Data mining, document analysis, chat-bots, Grammarly, document indexing / retrieval.



Applications

Music summarization

- **Input**: arbitrary-length audio.
- **Output**: shorter audio, preserving prologue+main+epilogue.
- **Applications**: Computer-aided music composition, music indexing, content-based music retrieval.



(Xu et. al. , 2004) Automatically Summarize Musical Audio Using Adaptive Clustering.

Applications

Video summarization

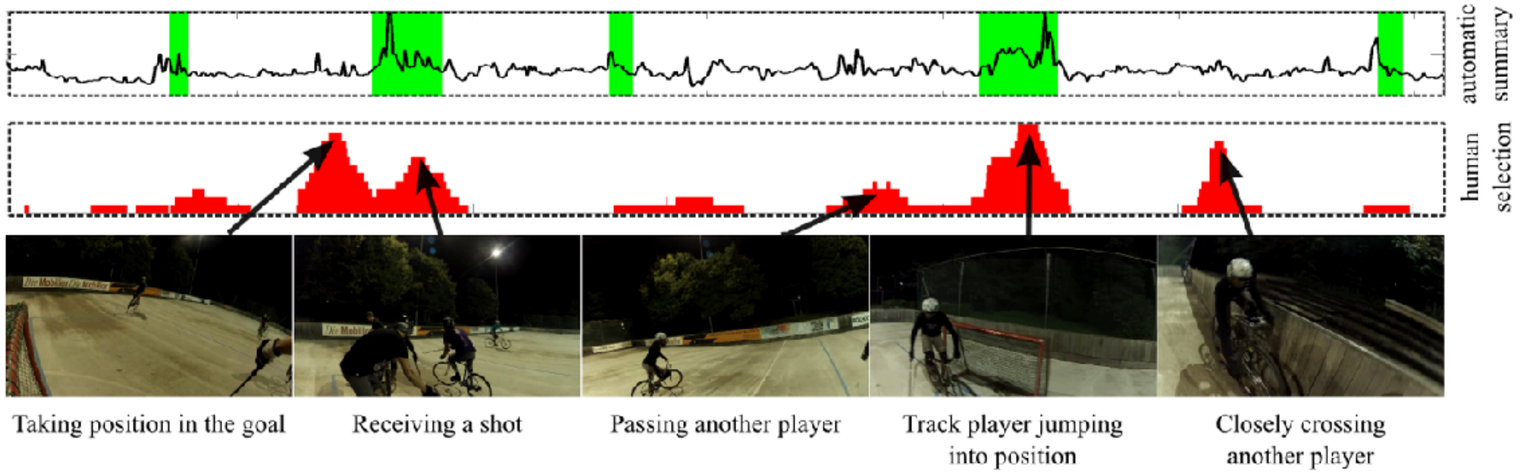
- **Input**: arbitrary-length video.
- **Output**: shorter video, preserving main and most important scenes.
- **Applications**: Video analysis, video indexing, content-based video retrieval.

(Gigly et.al. , 2014) Creating Summaries from User Videos.

Applications

Video summarization

- **Input**: arbitrary-length video.
- **Output**: shorter video, preserving main and most important scenes.
- **Applications**: Video analysis, video indexing, content-based video retrieval.



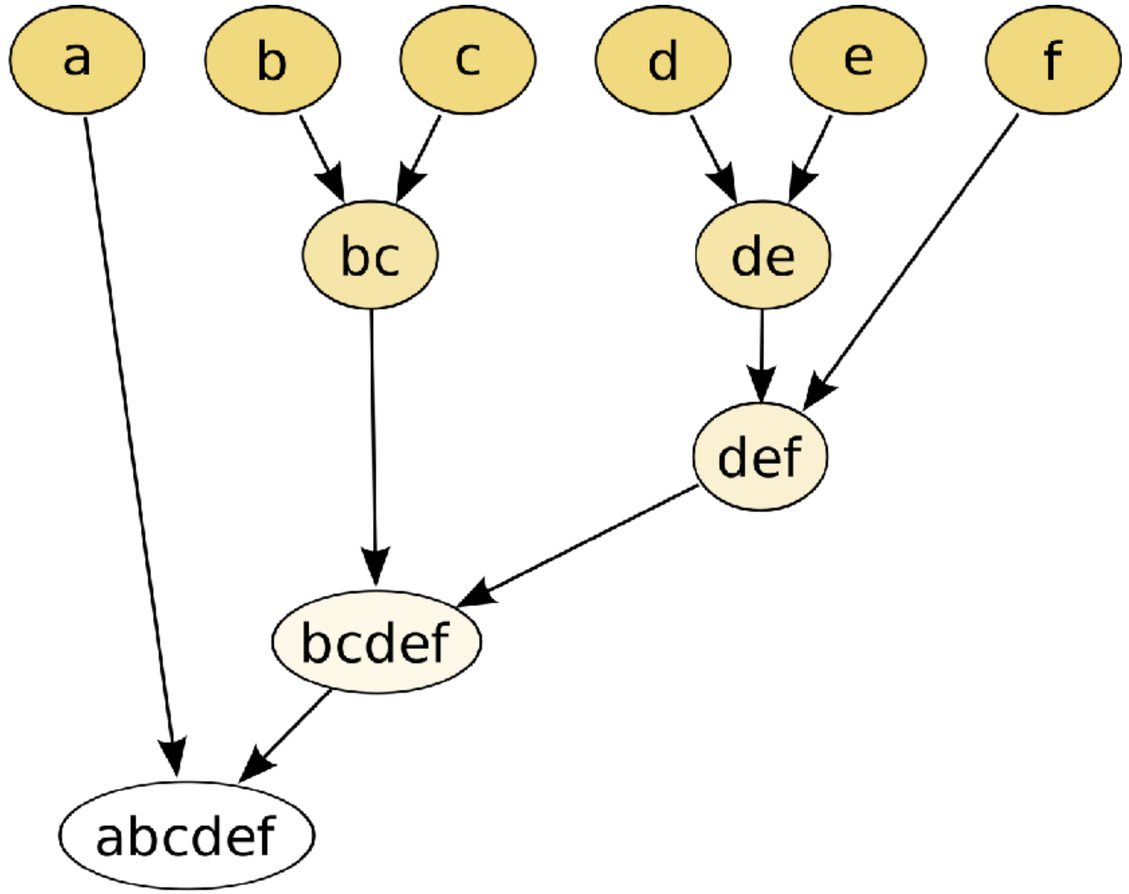
(Gigly et.al. , 2014) Creating Summaries from User Videos.

3. Hierarchical Clustering

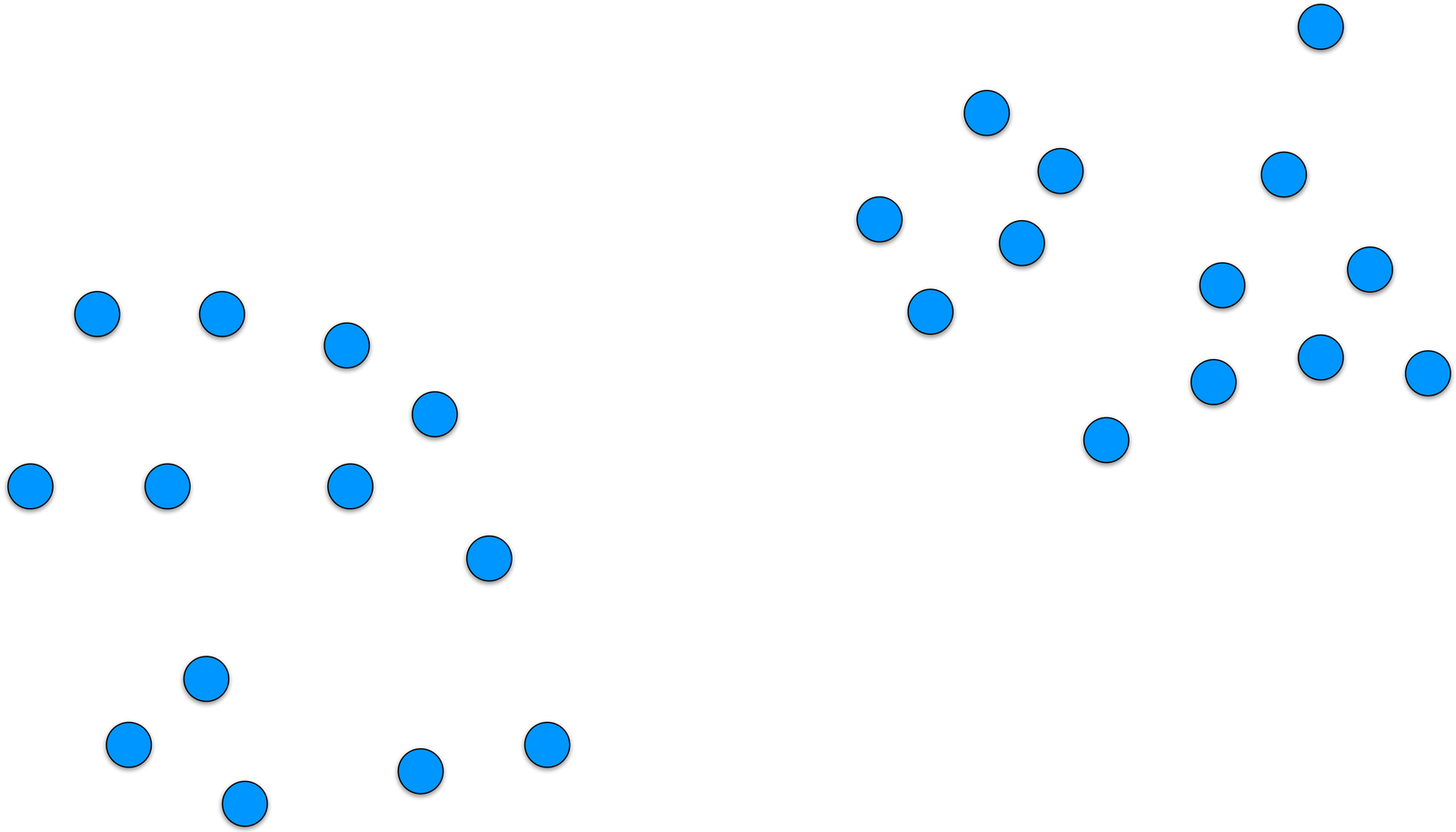
Hierarchical Clustering

General Idea:

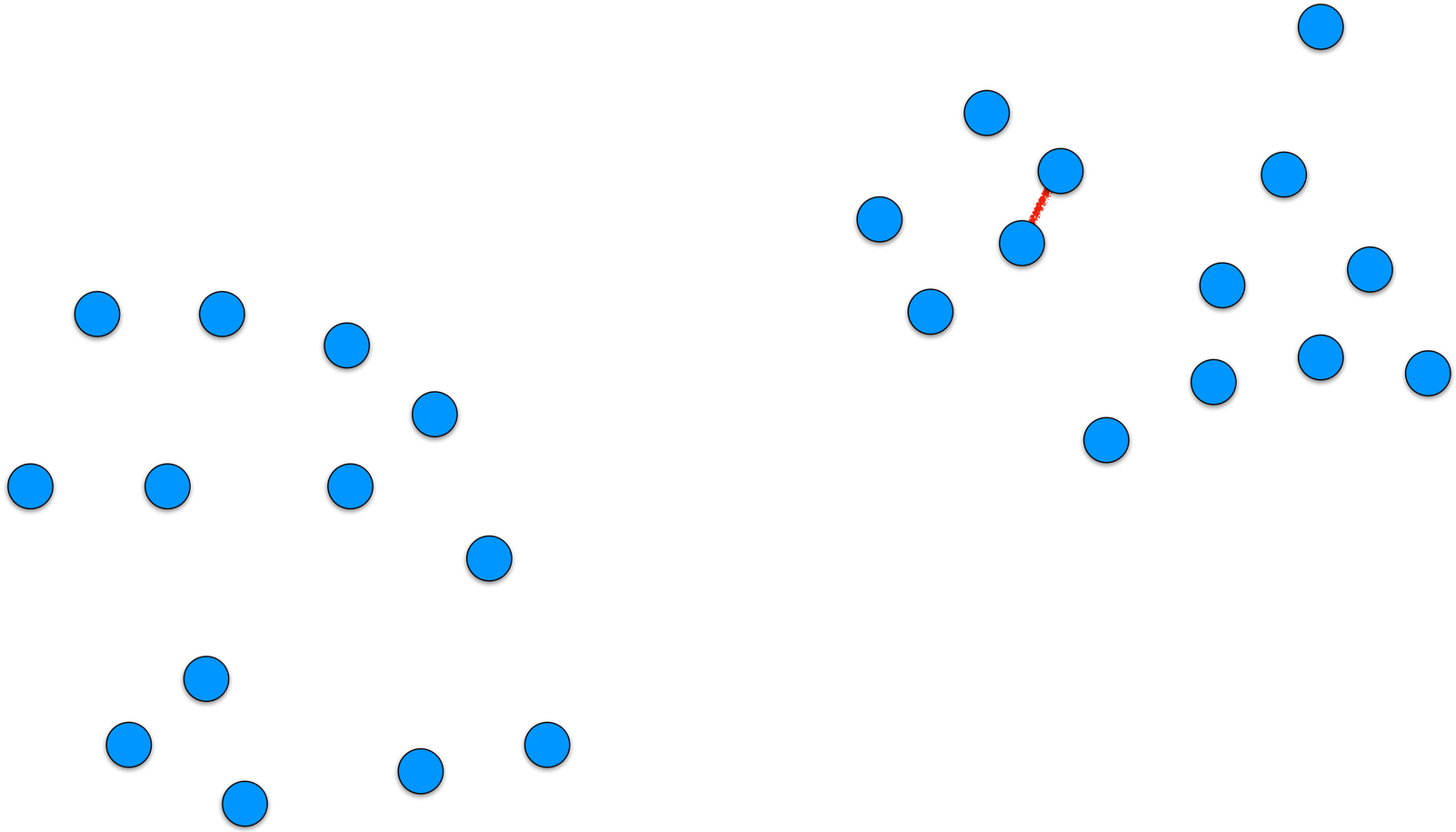
- Find the closest two points and cluster them
- The clustered points are considered as a new point
- Repeat until the distance of the closest two points is too large



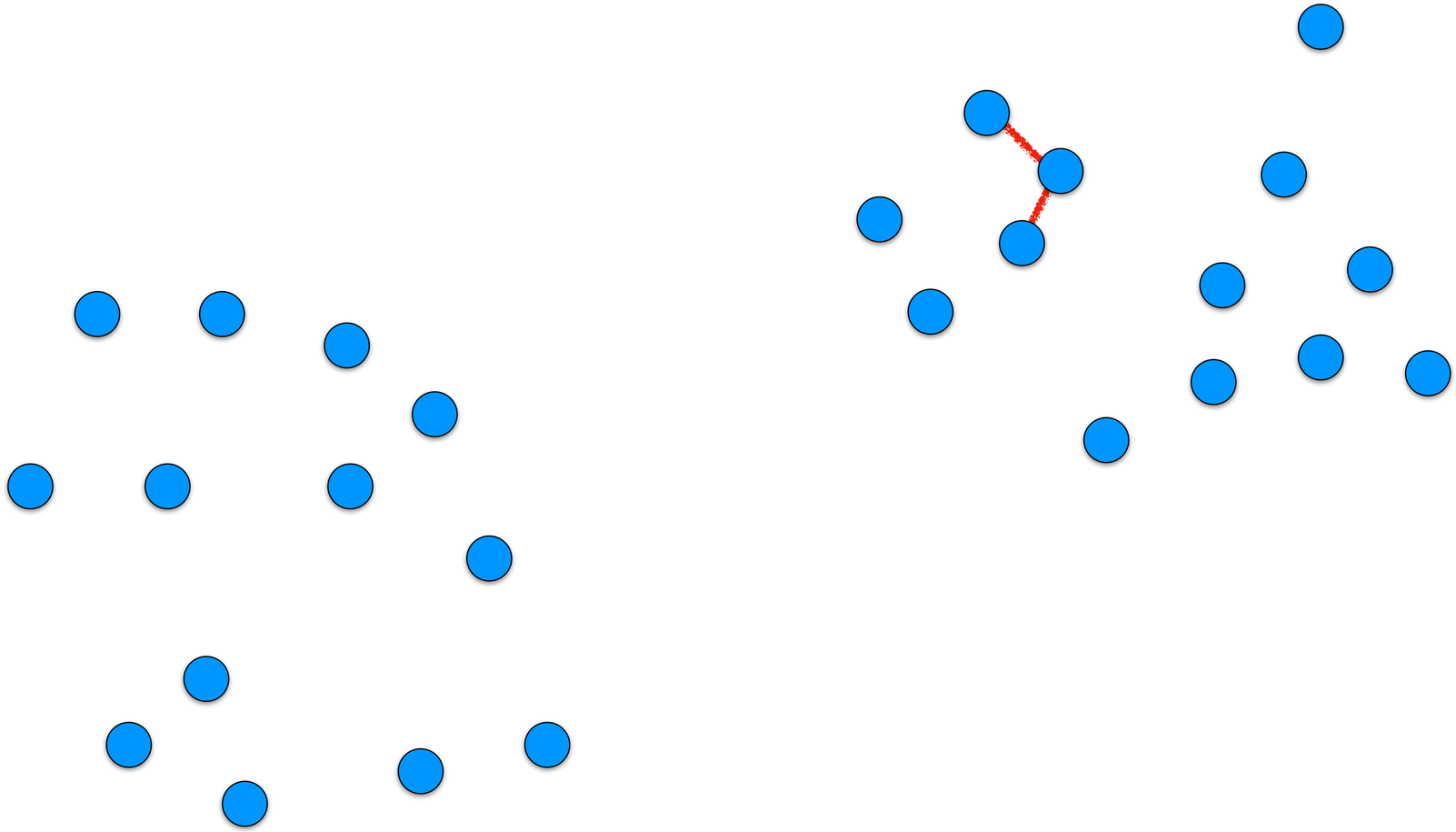
Algorithm



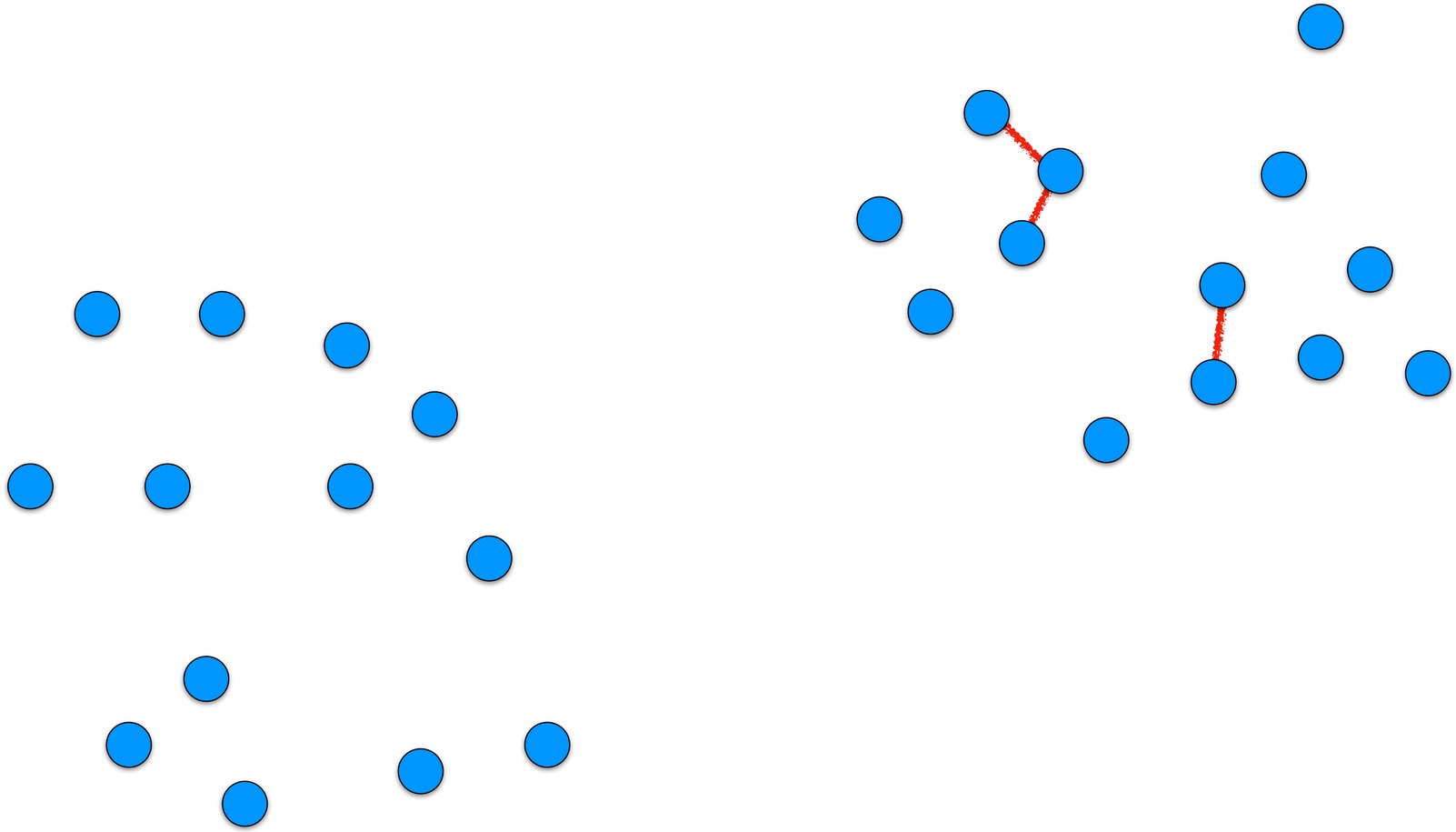
Algorithm



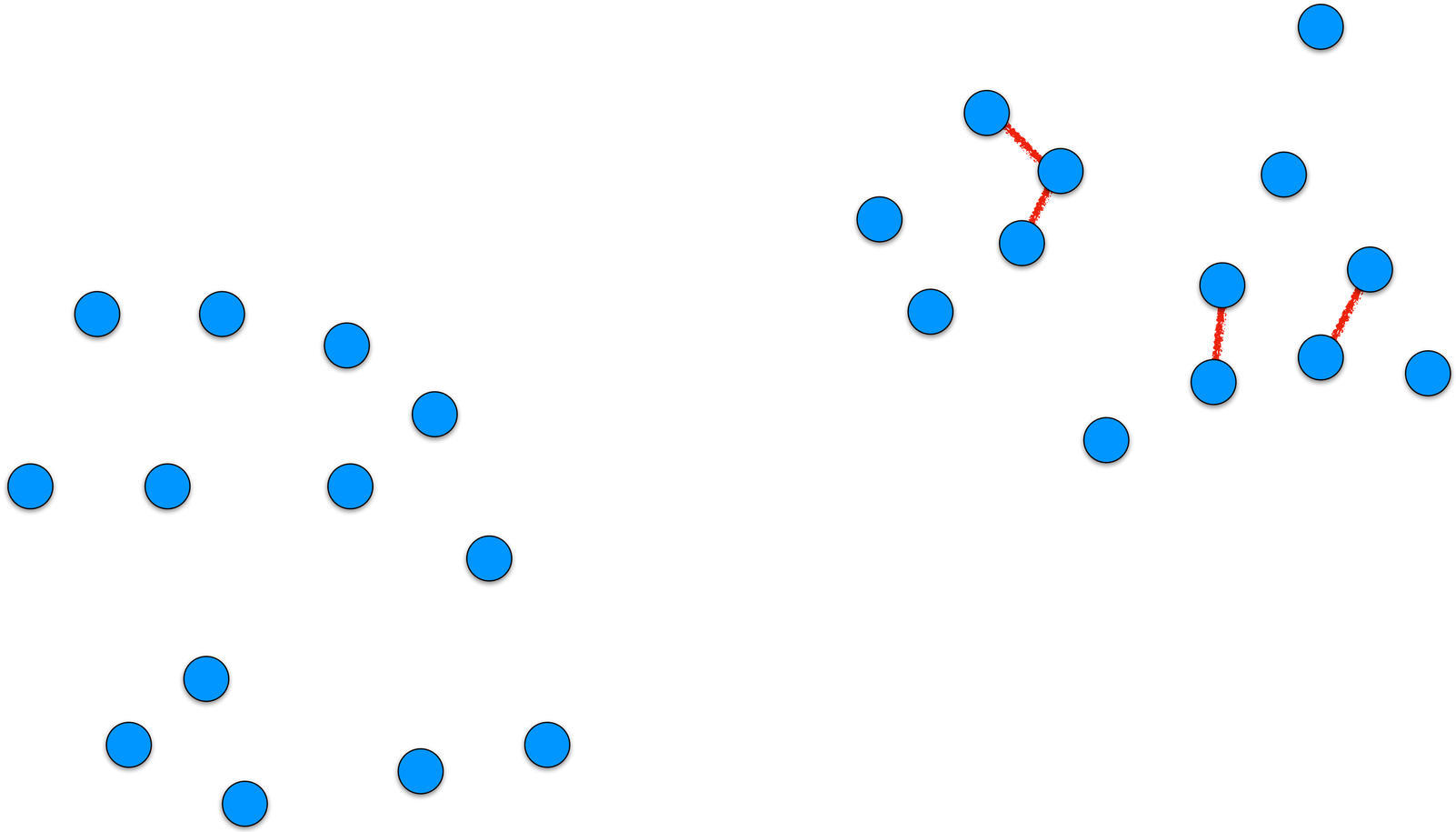
Algorithm



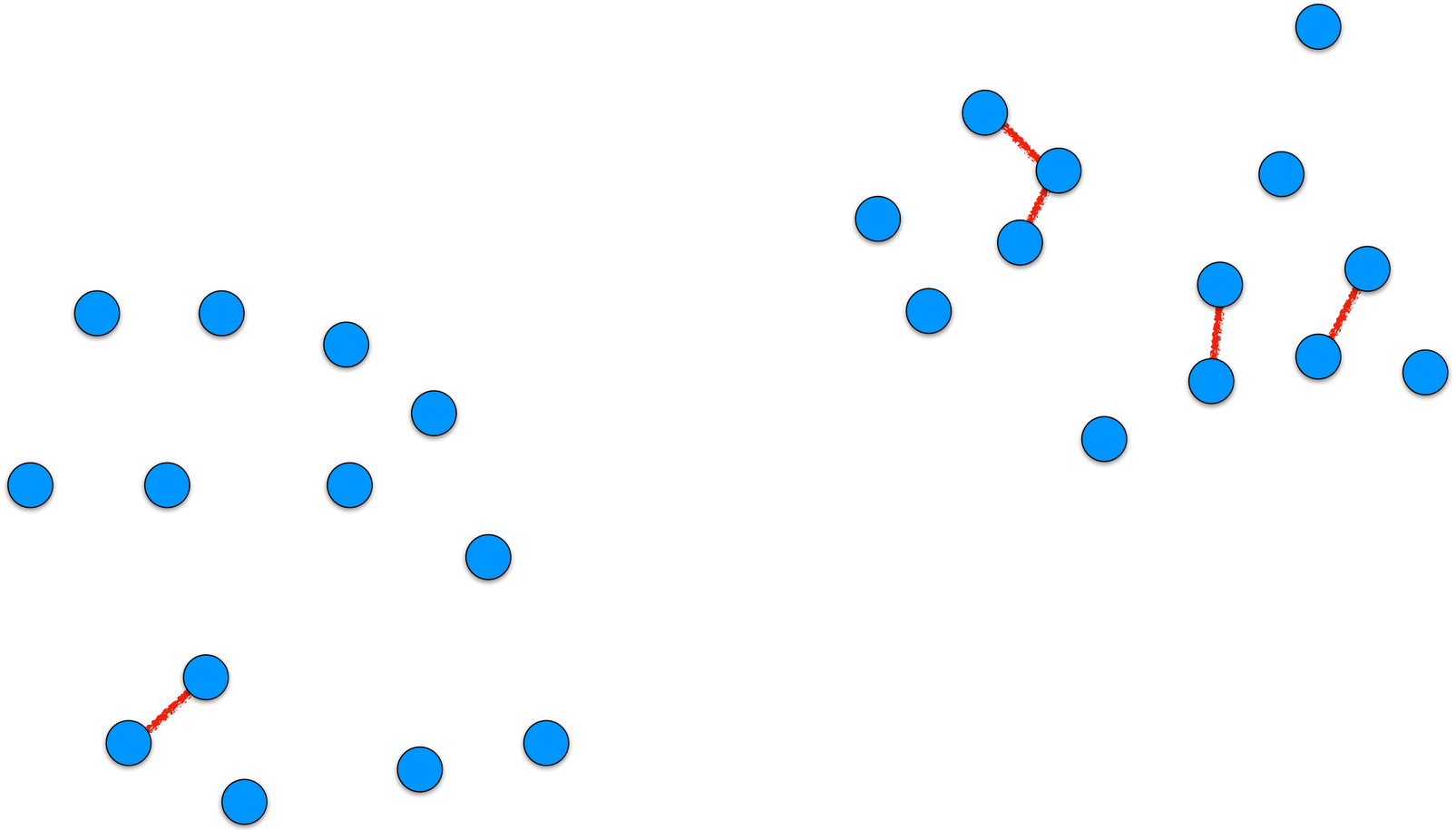
Algorithm



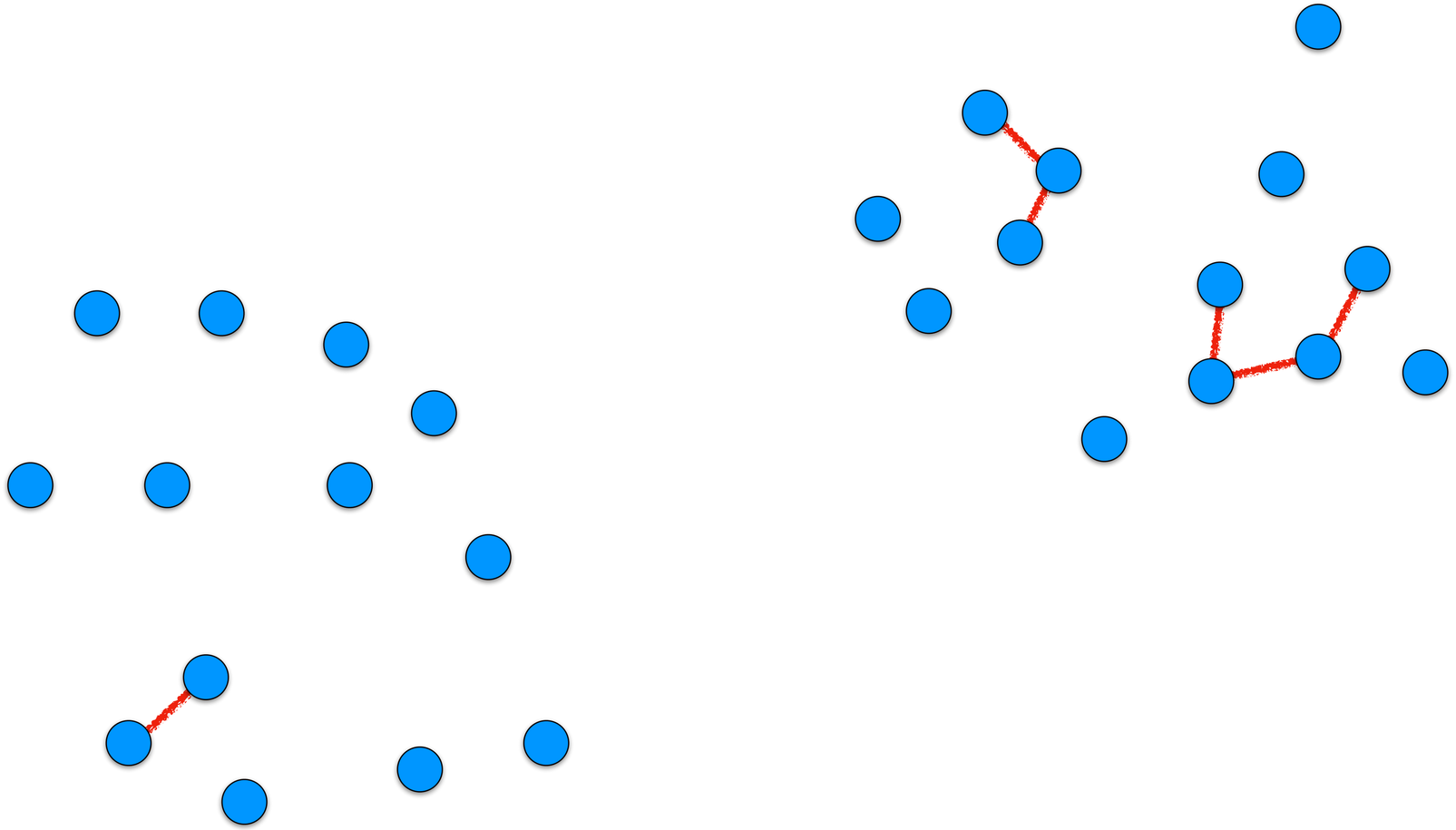
Algorithm



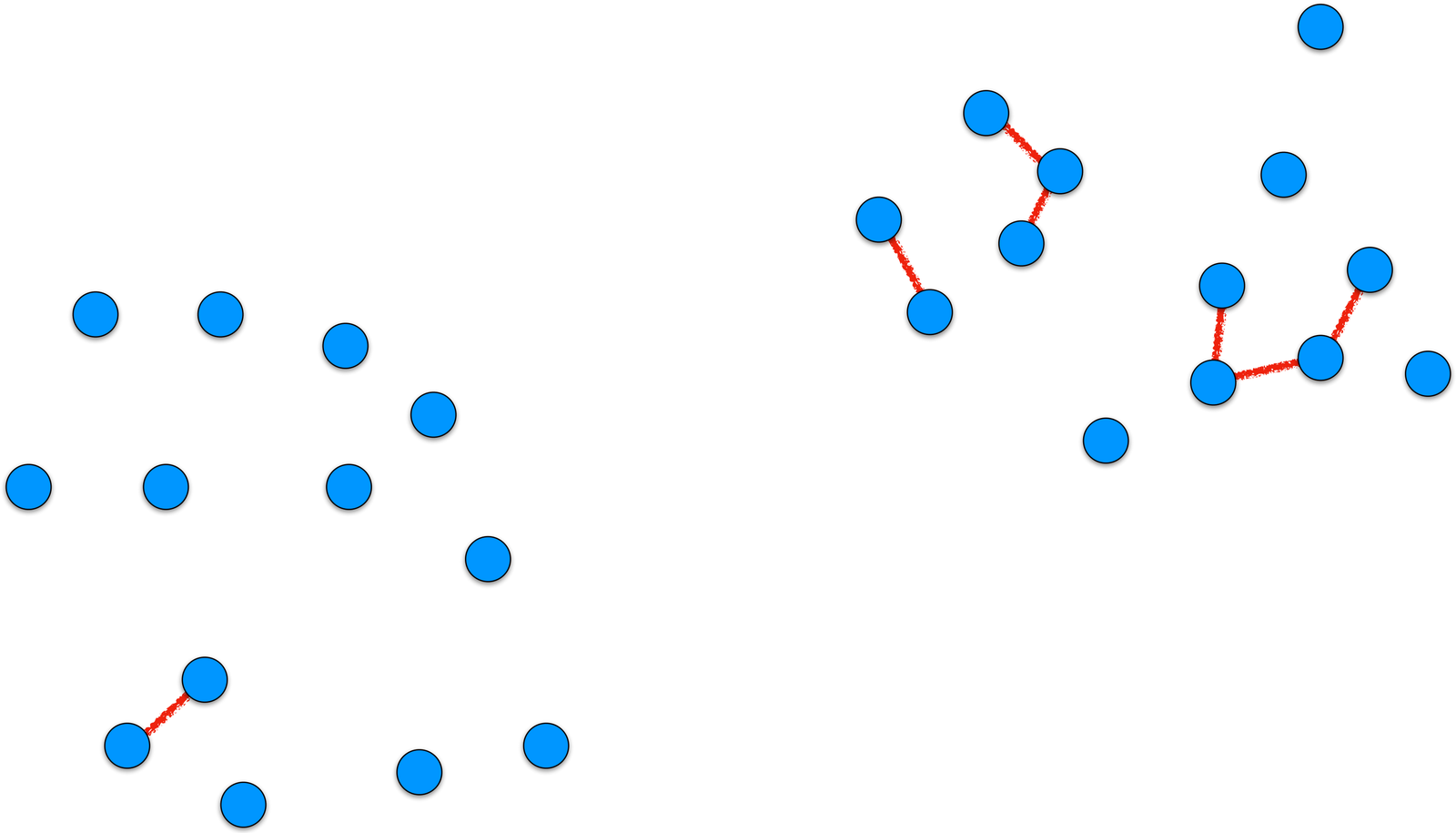
Algorithm



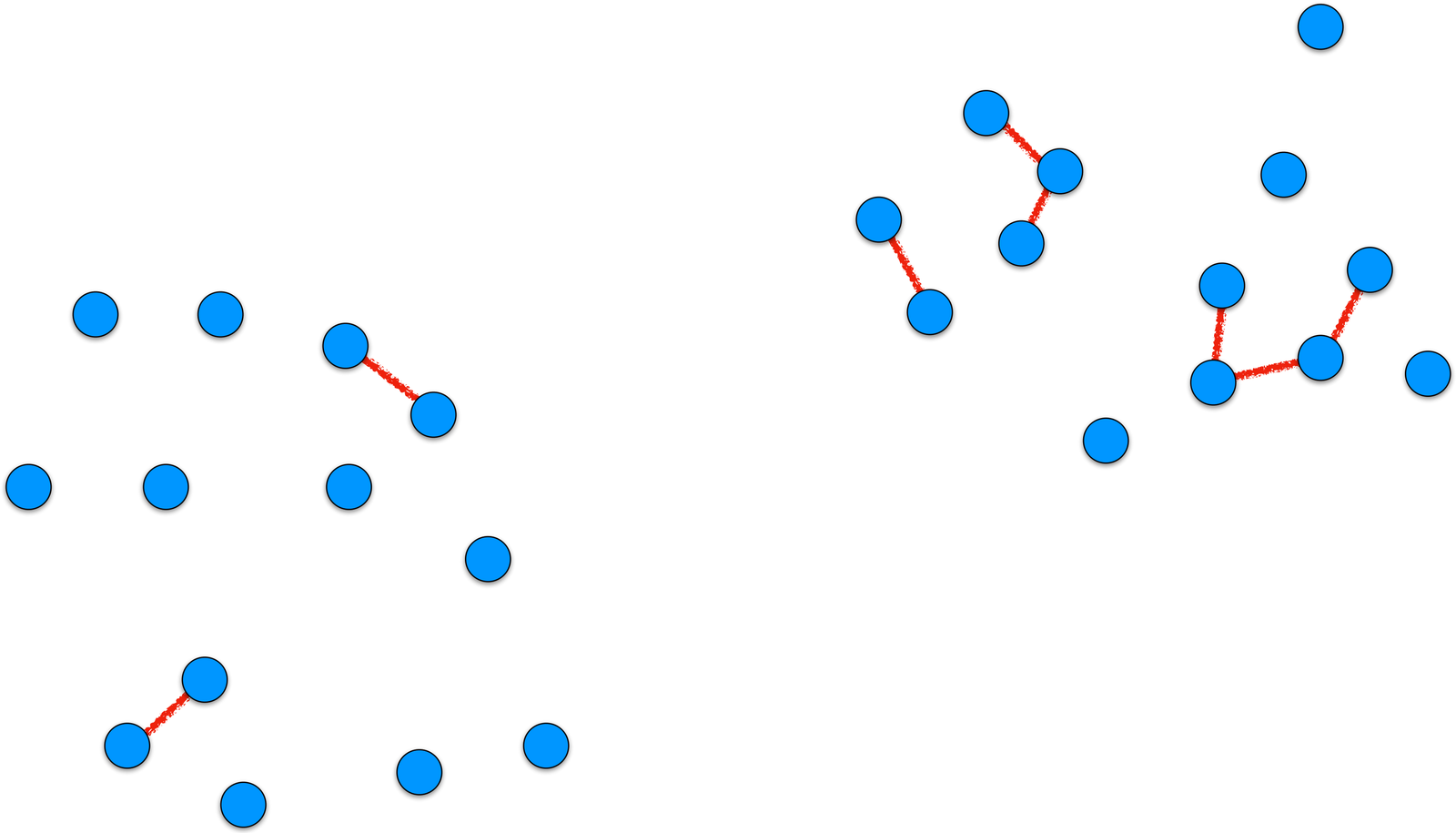
Algorithm



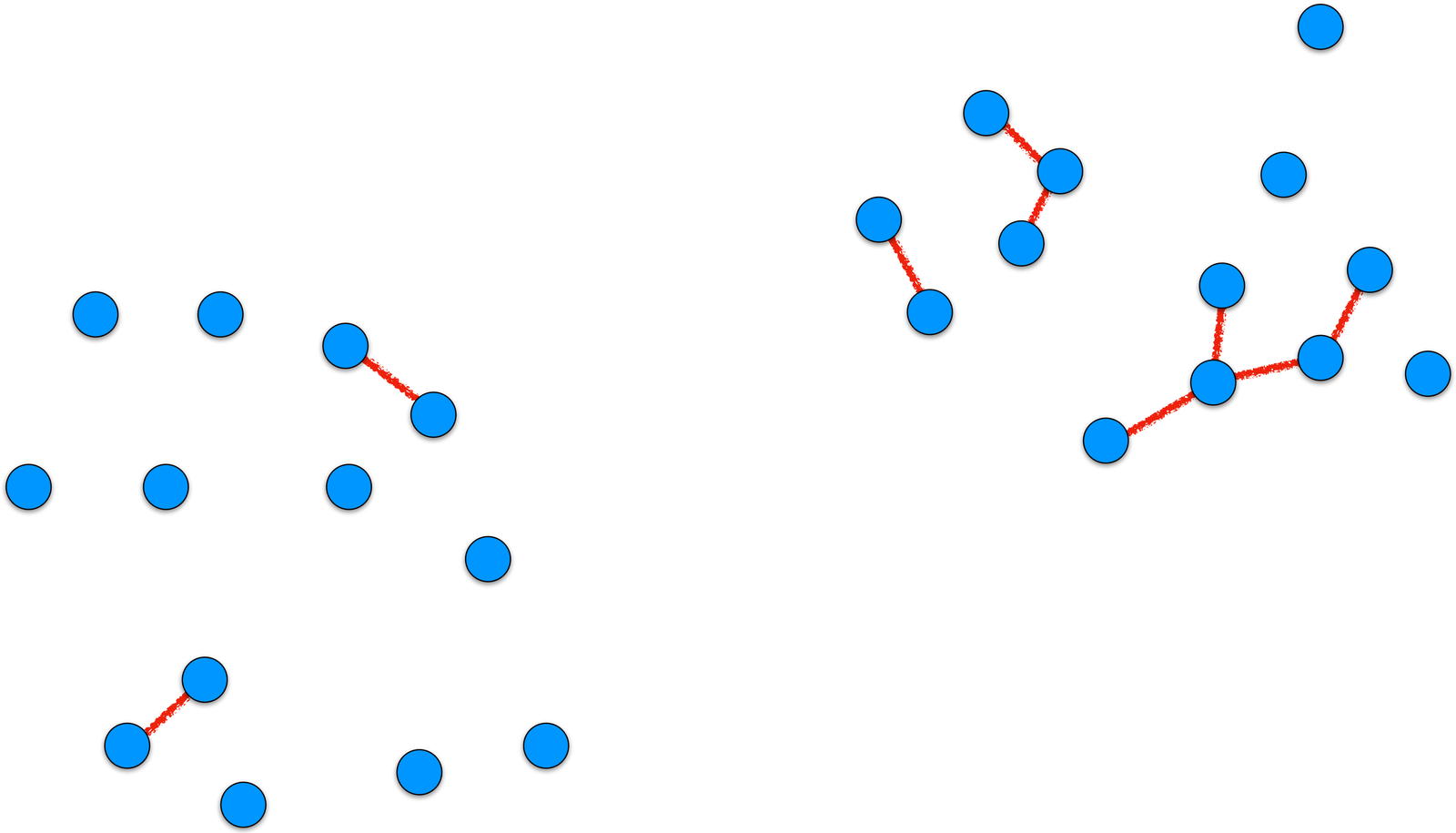
Algorithm



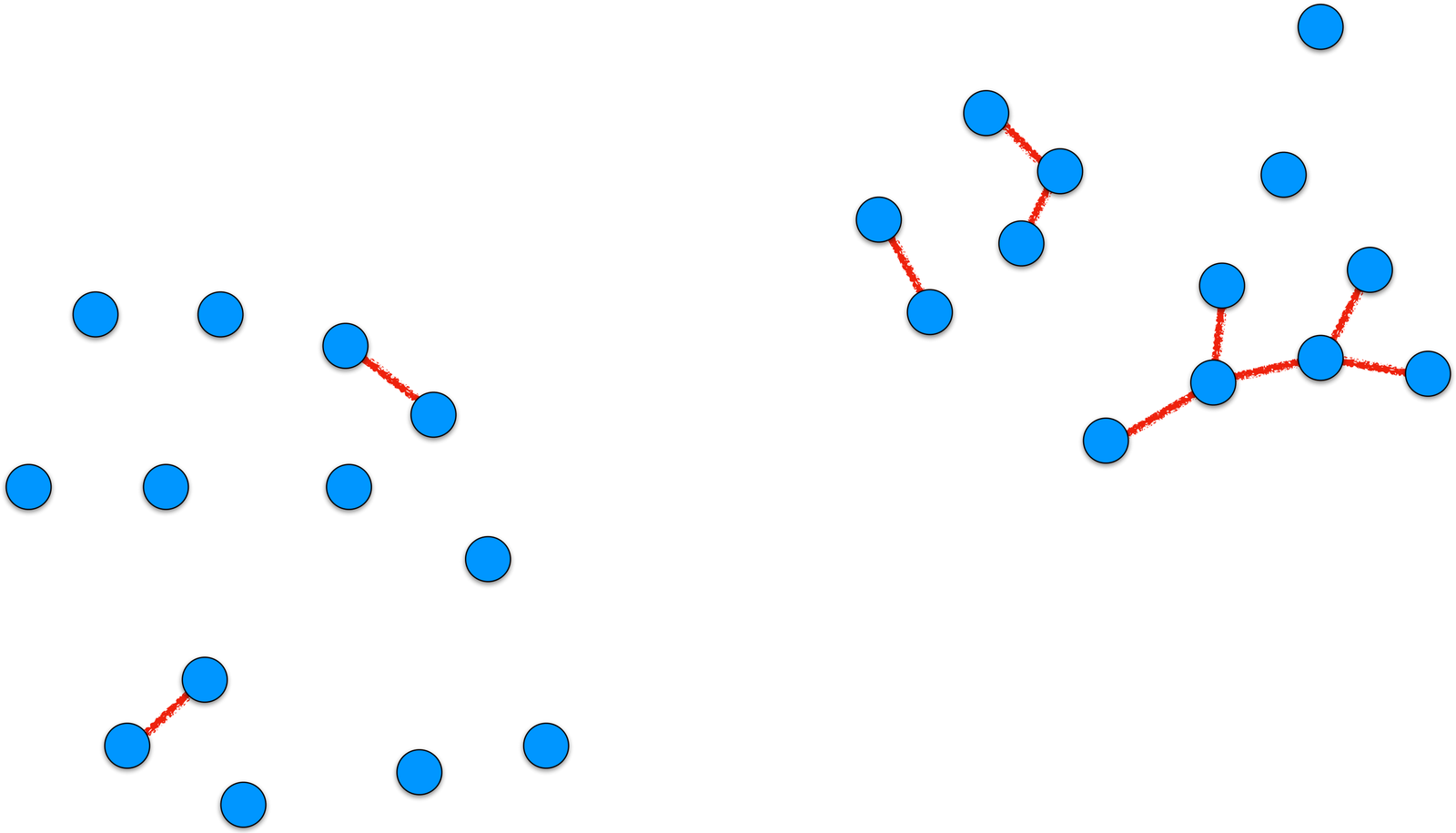
Algorithm



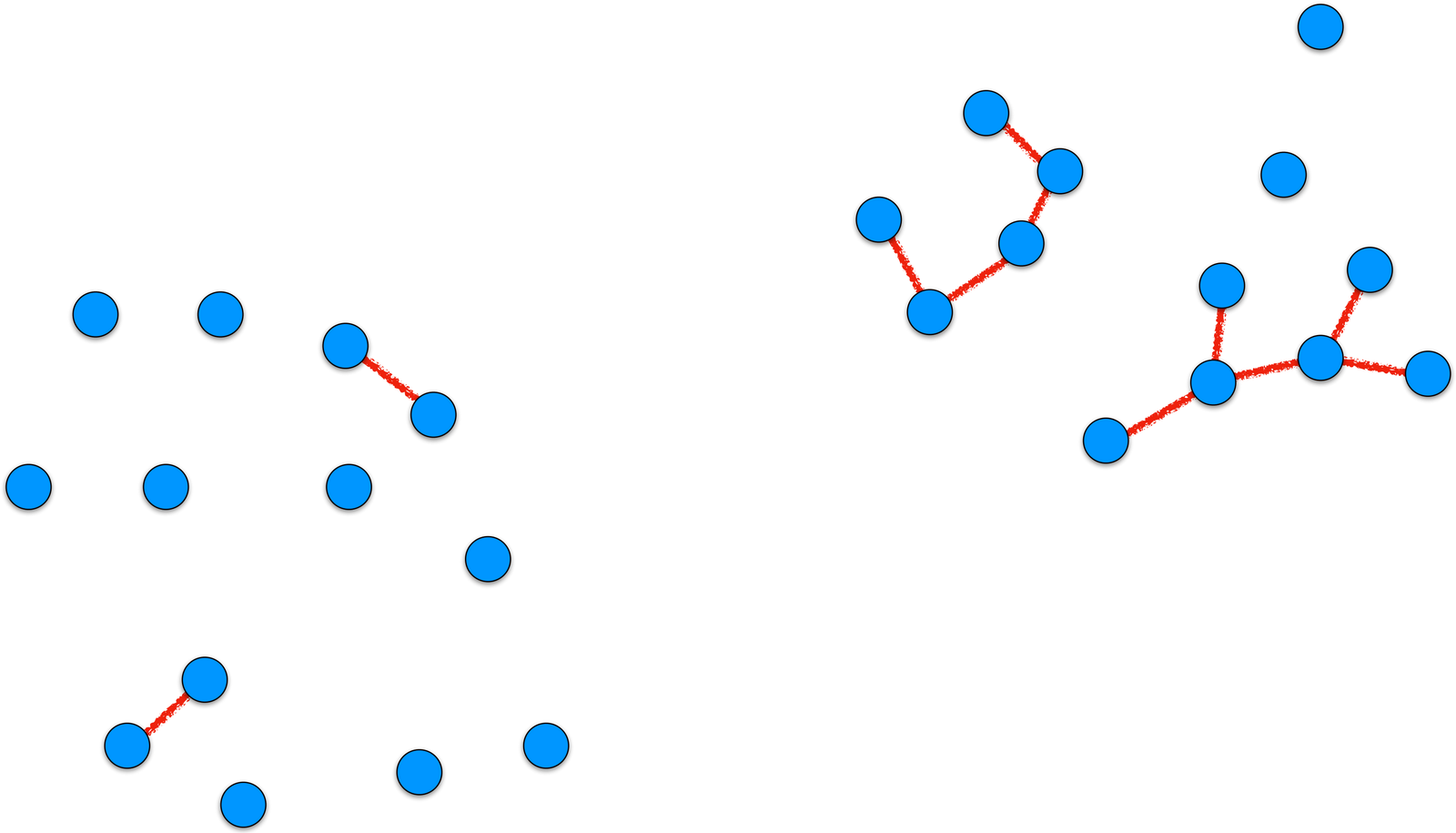
Algorithm



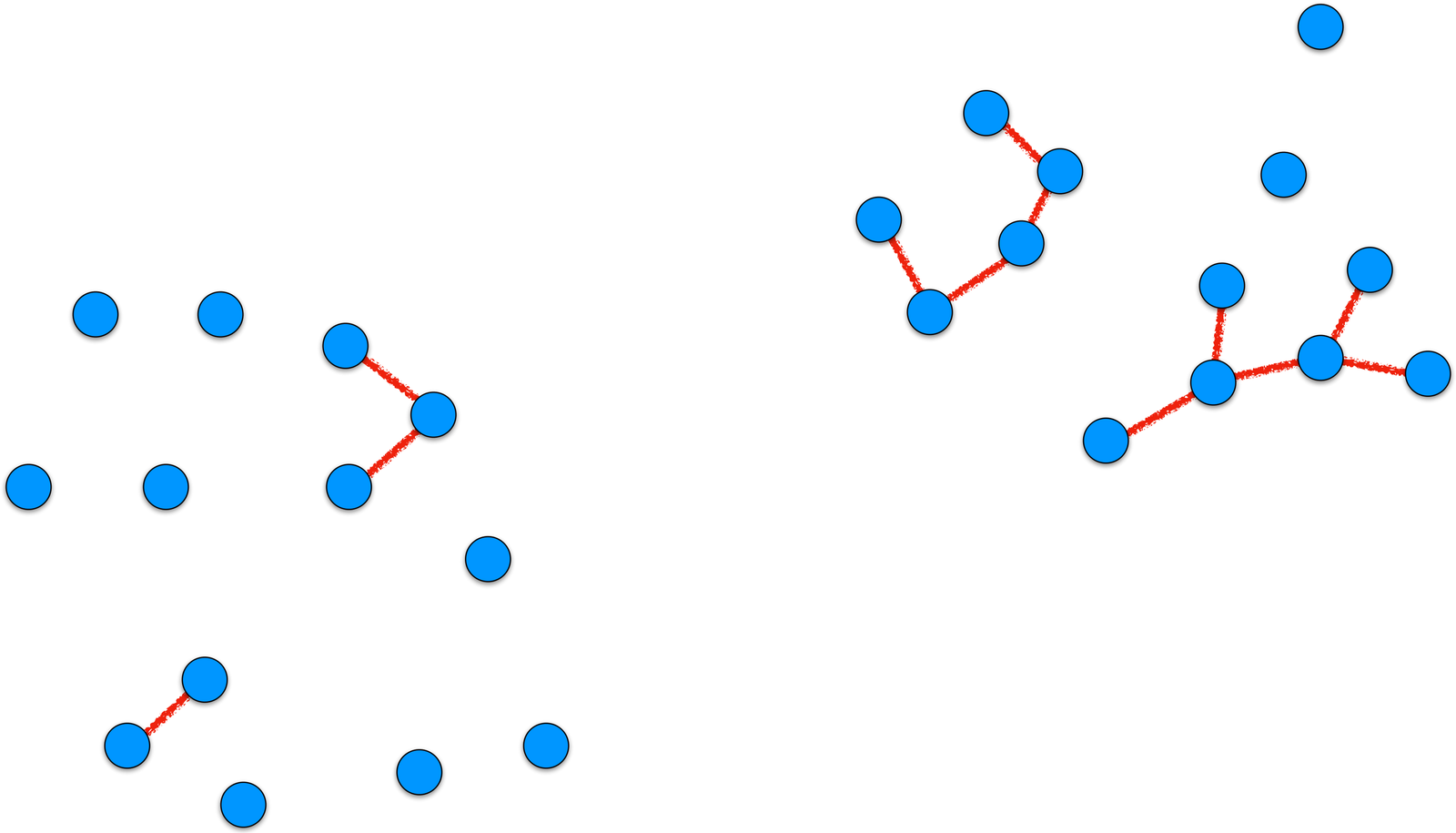
Algorithm



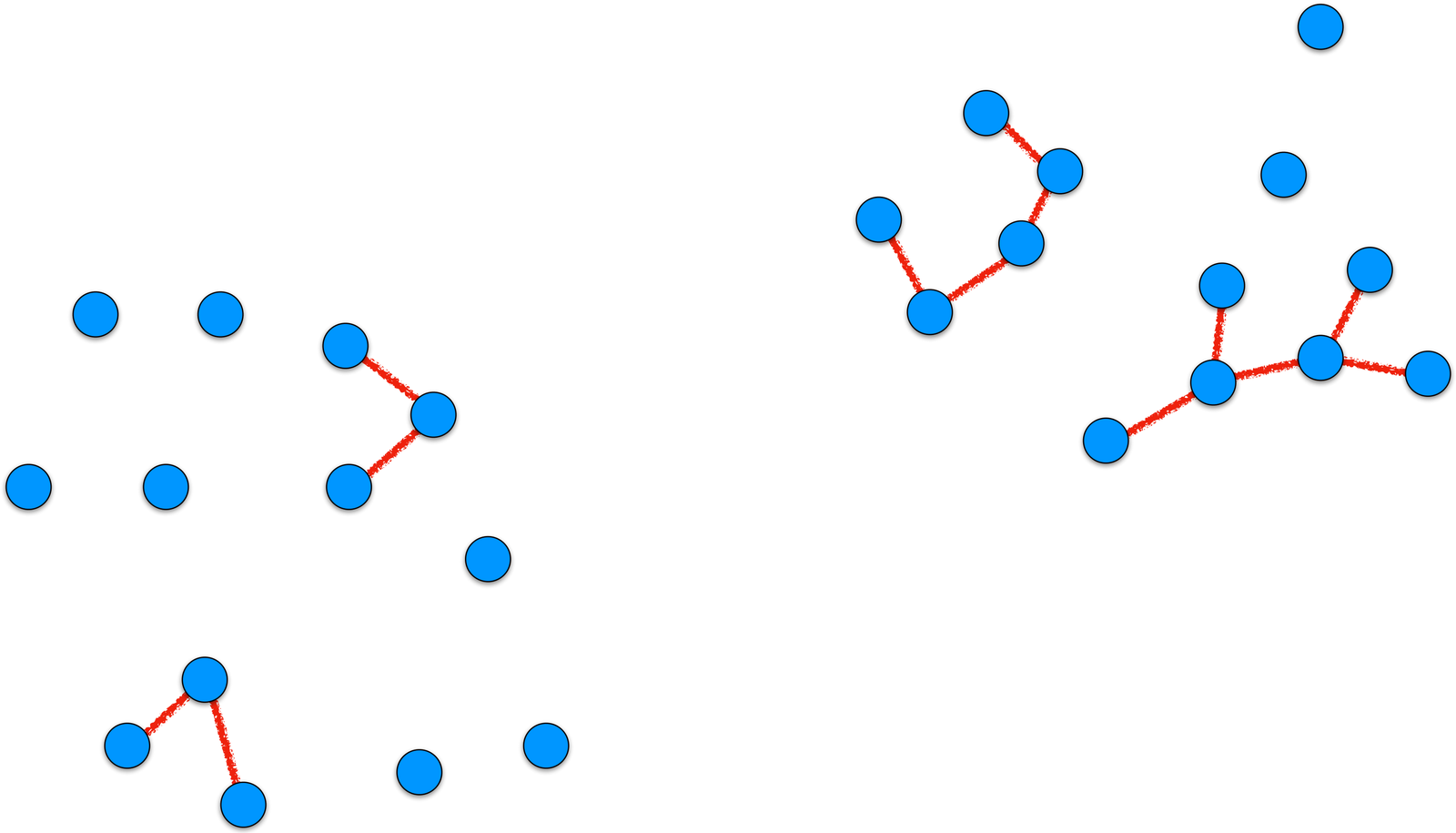
Algorithm



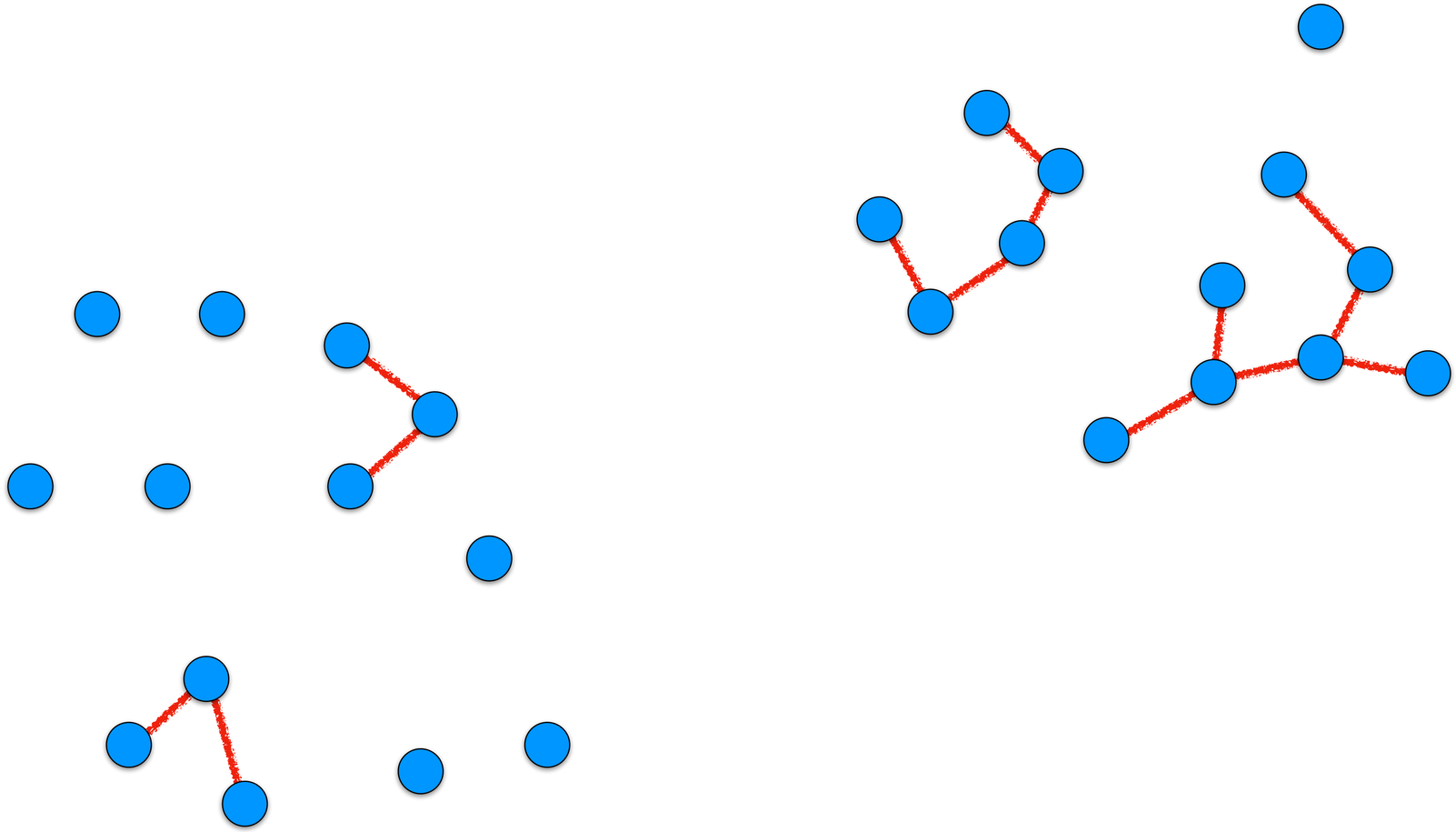
Algorithm



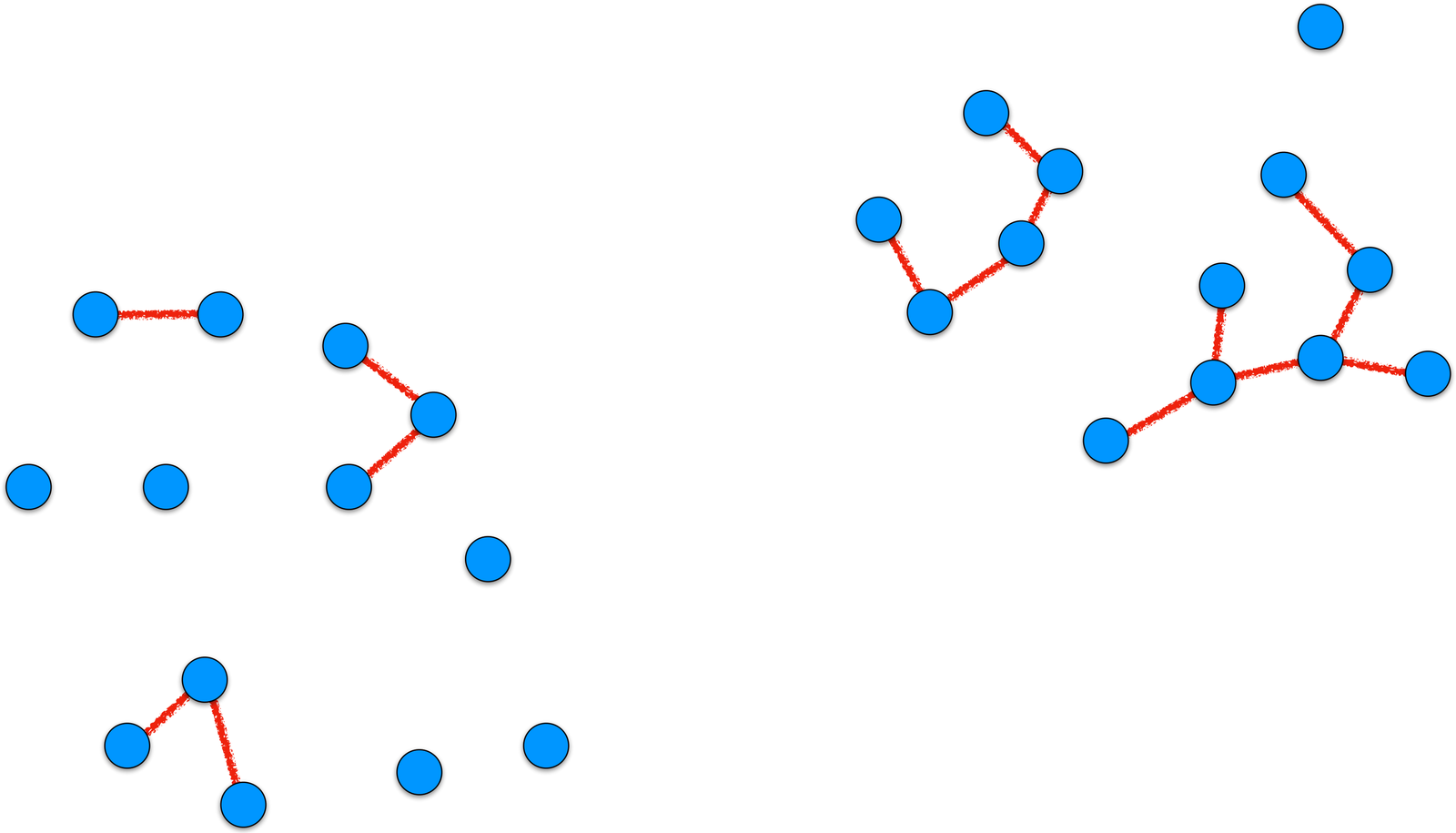
Algorithm



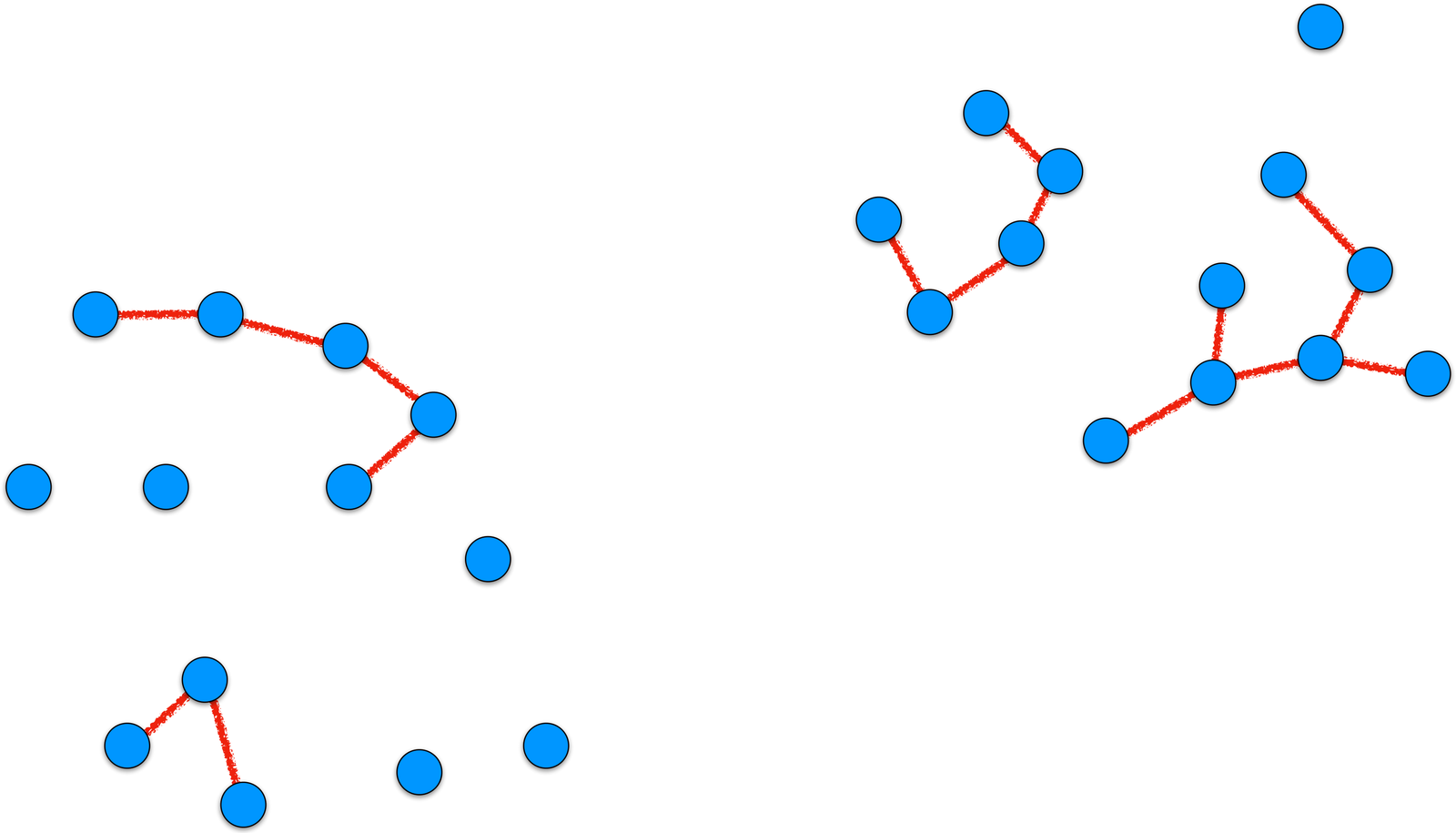
Algorithm



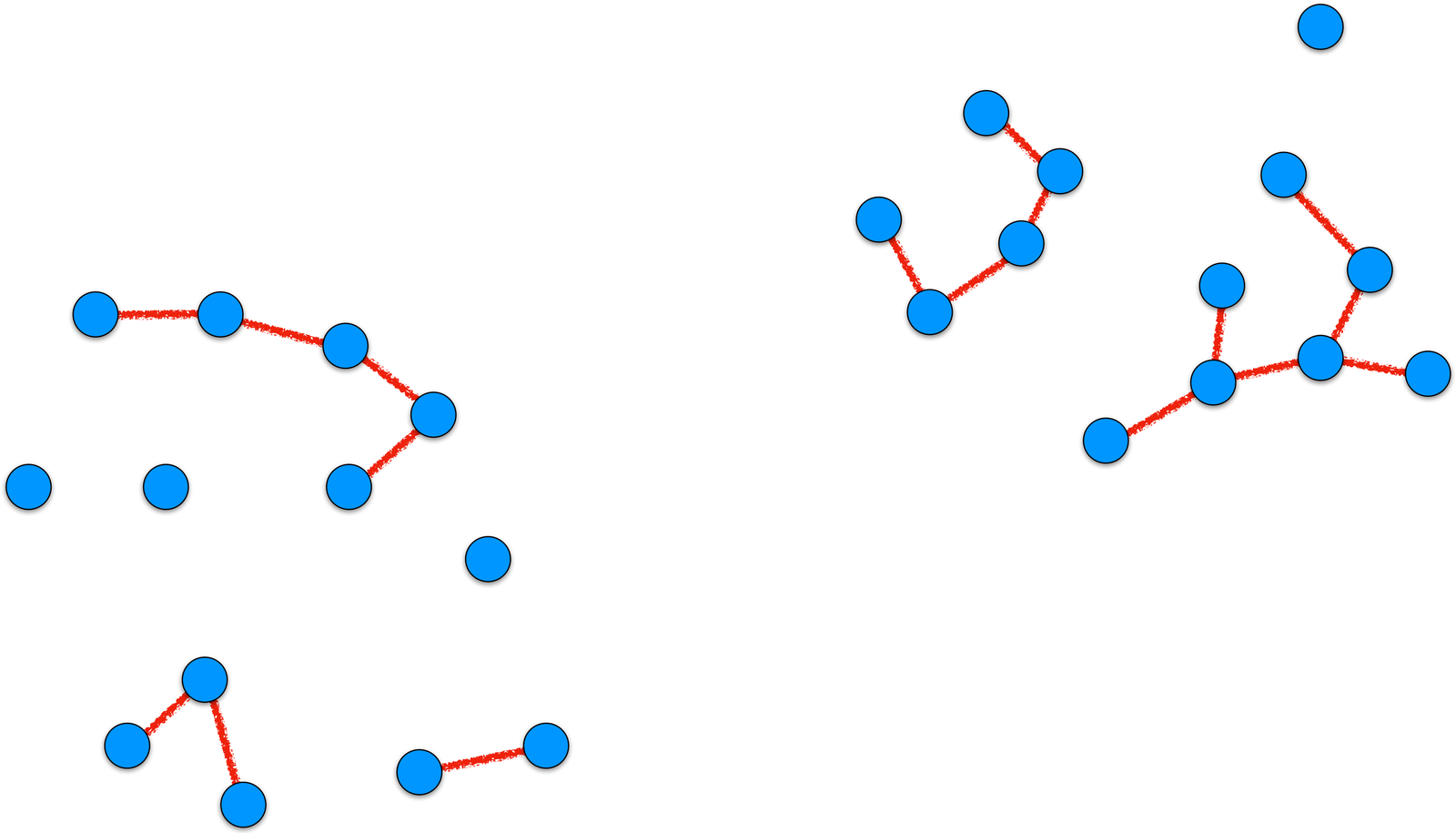
Algorithm



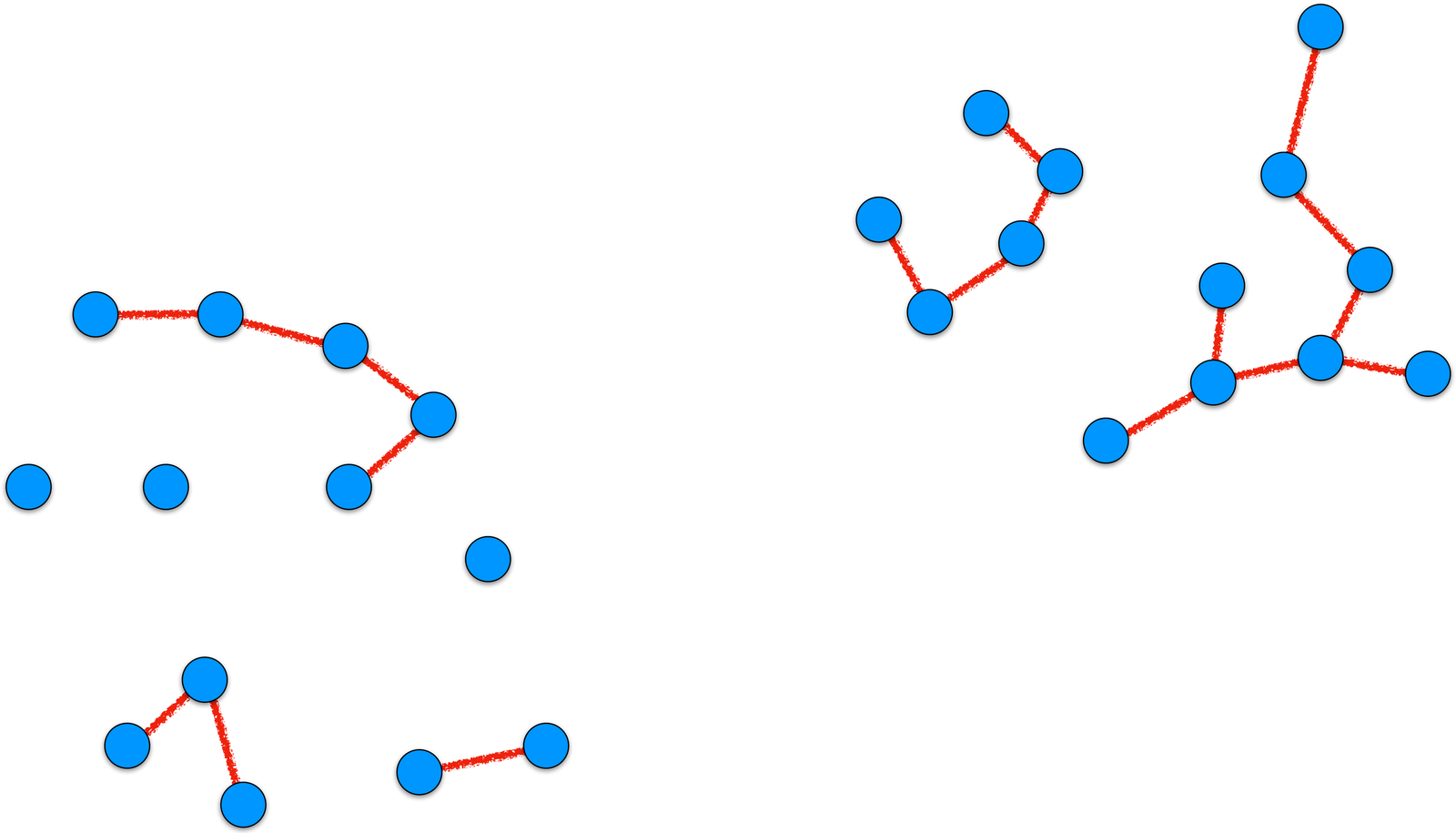
Algorithm



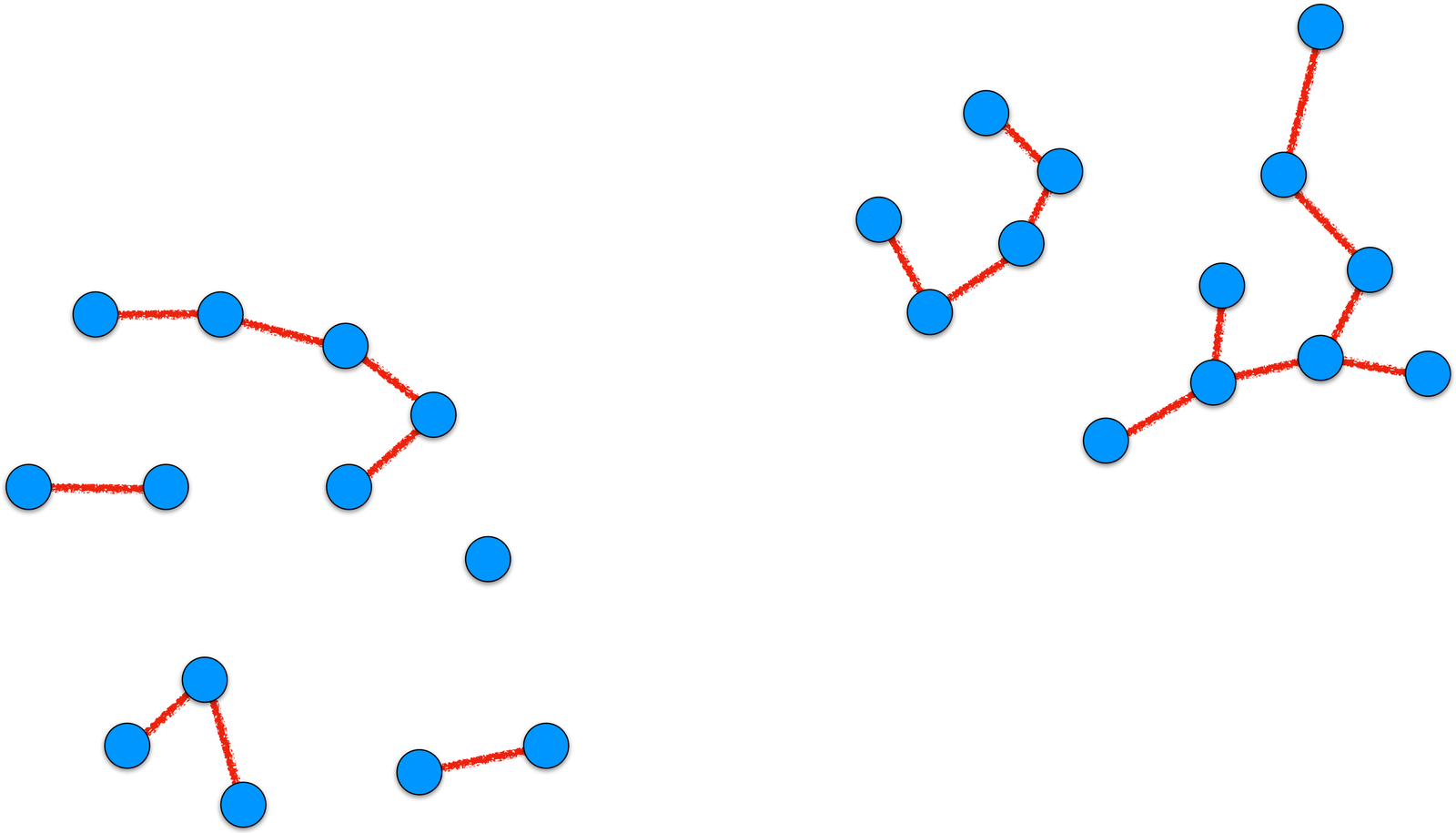
Algorithm



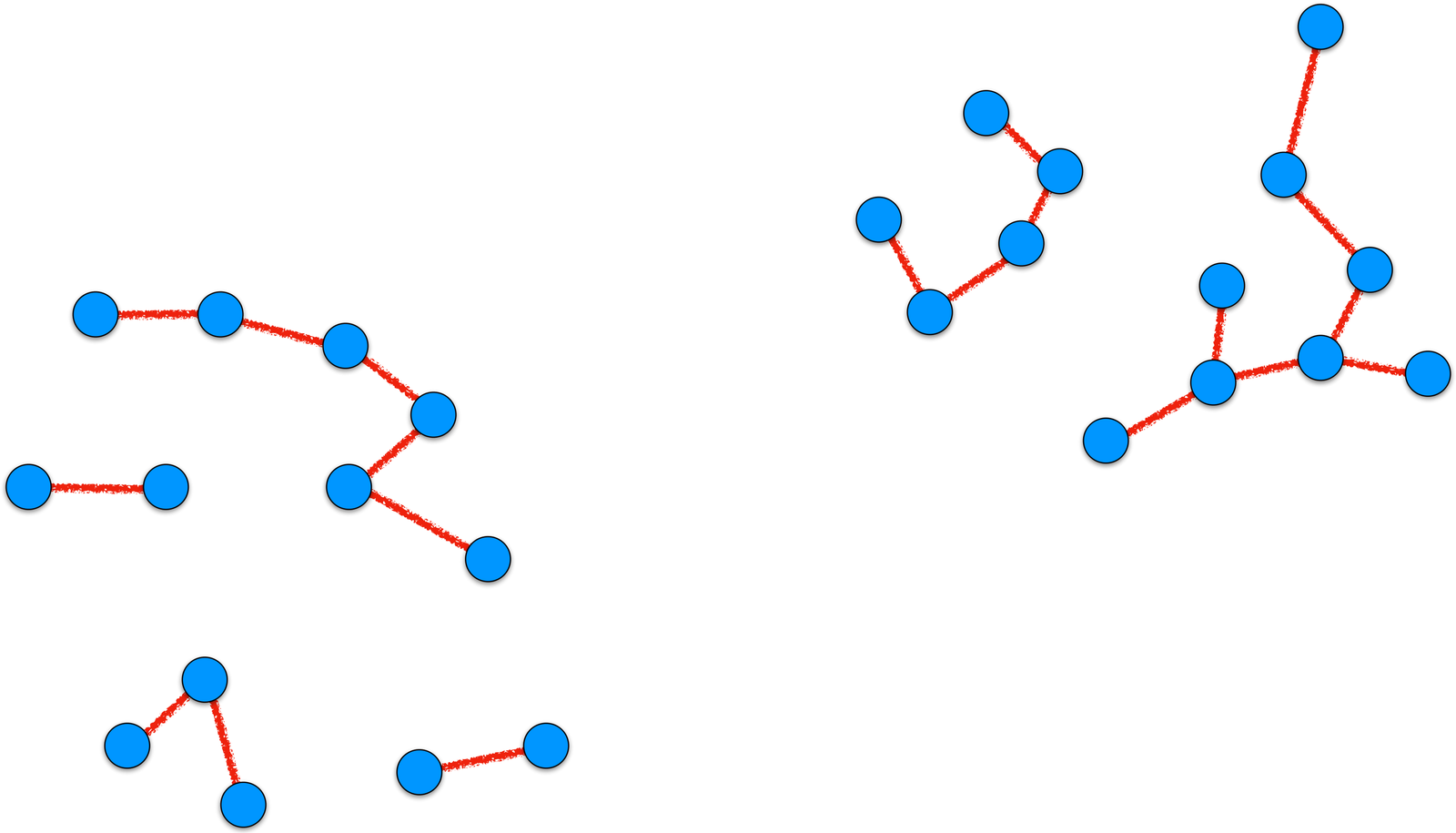
Algorithm



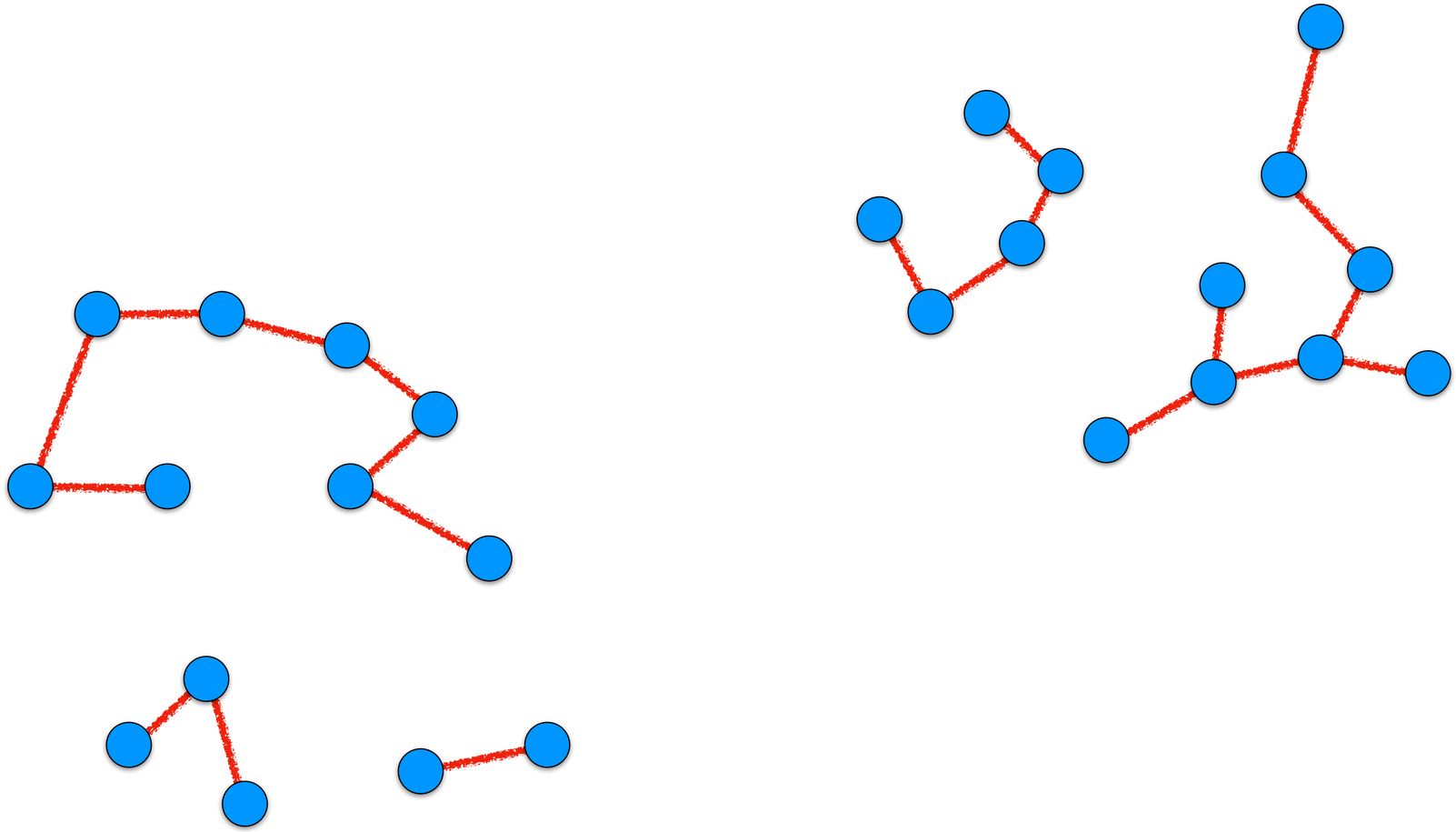
Algorithm



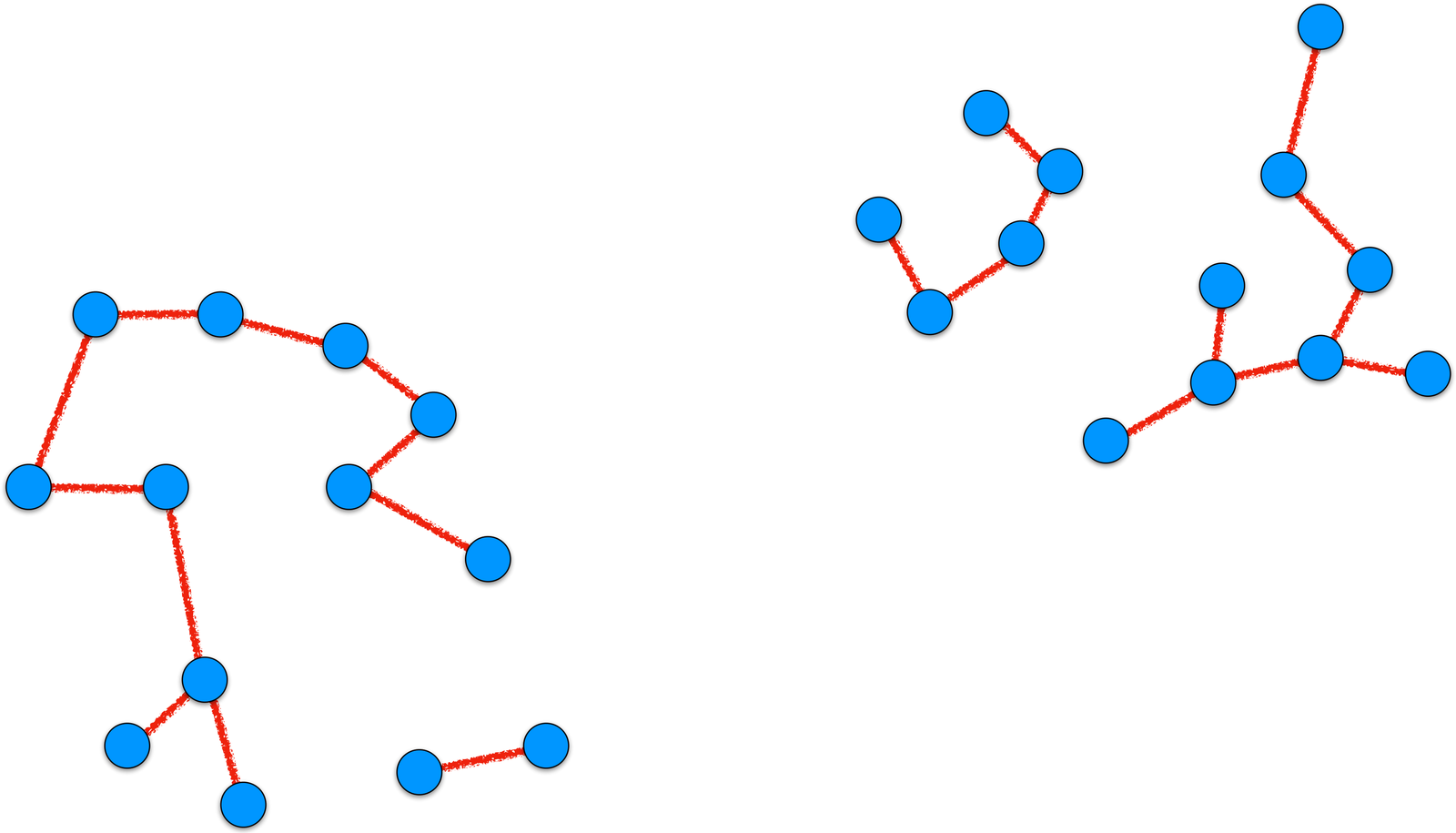
Algorithm



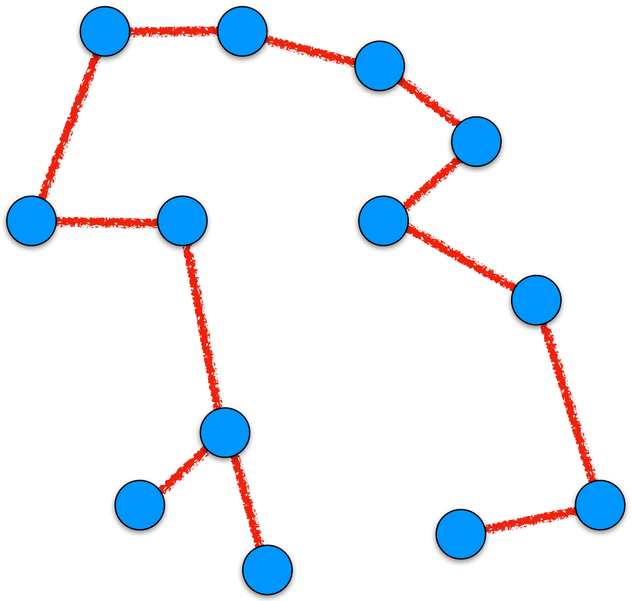
Algorithm

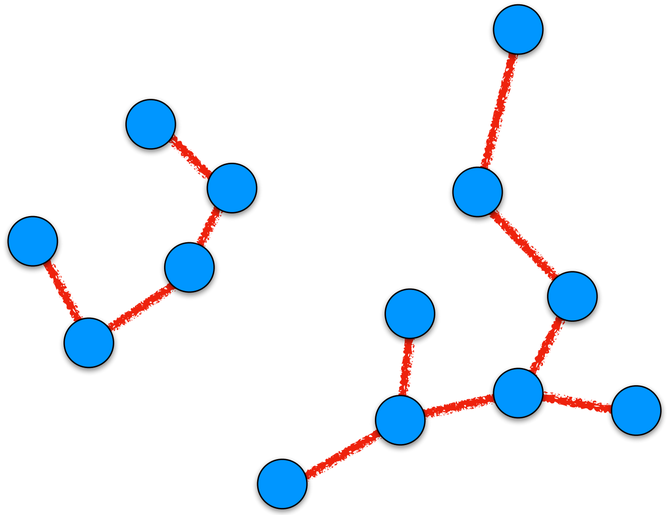


Algorithm

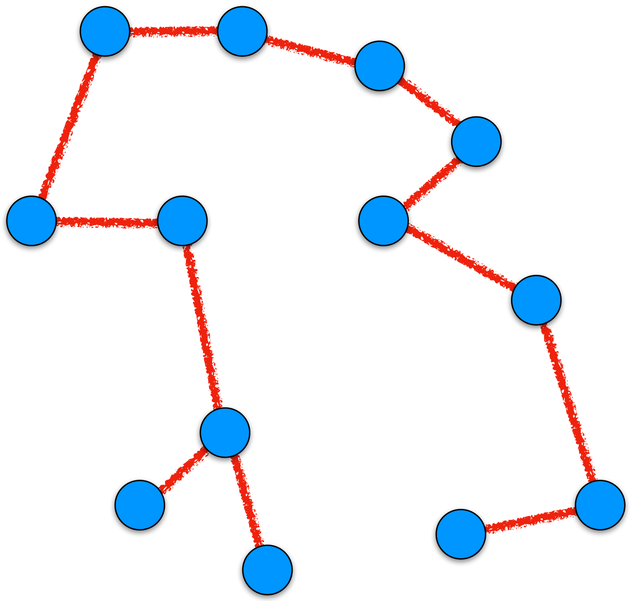


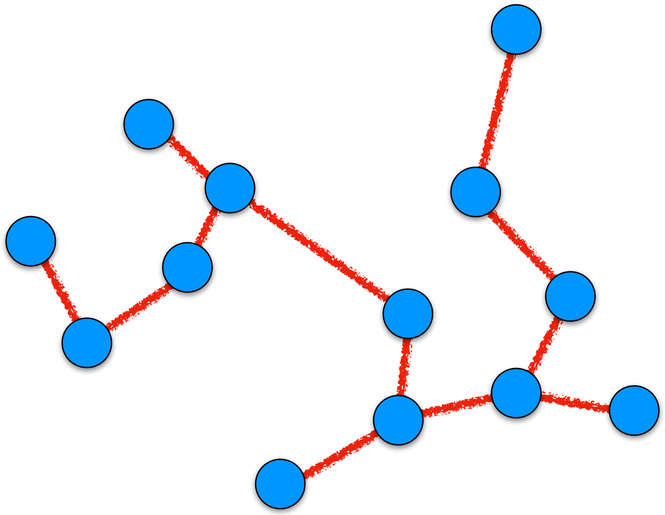
Algorithm



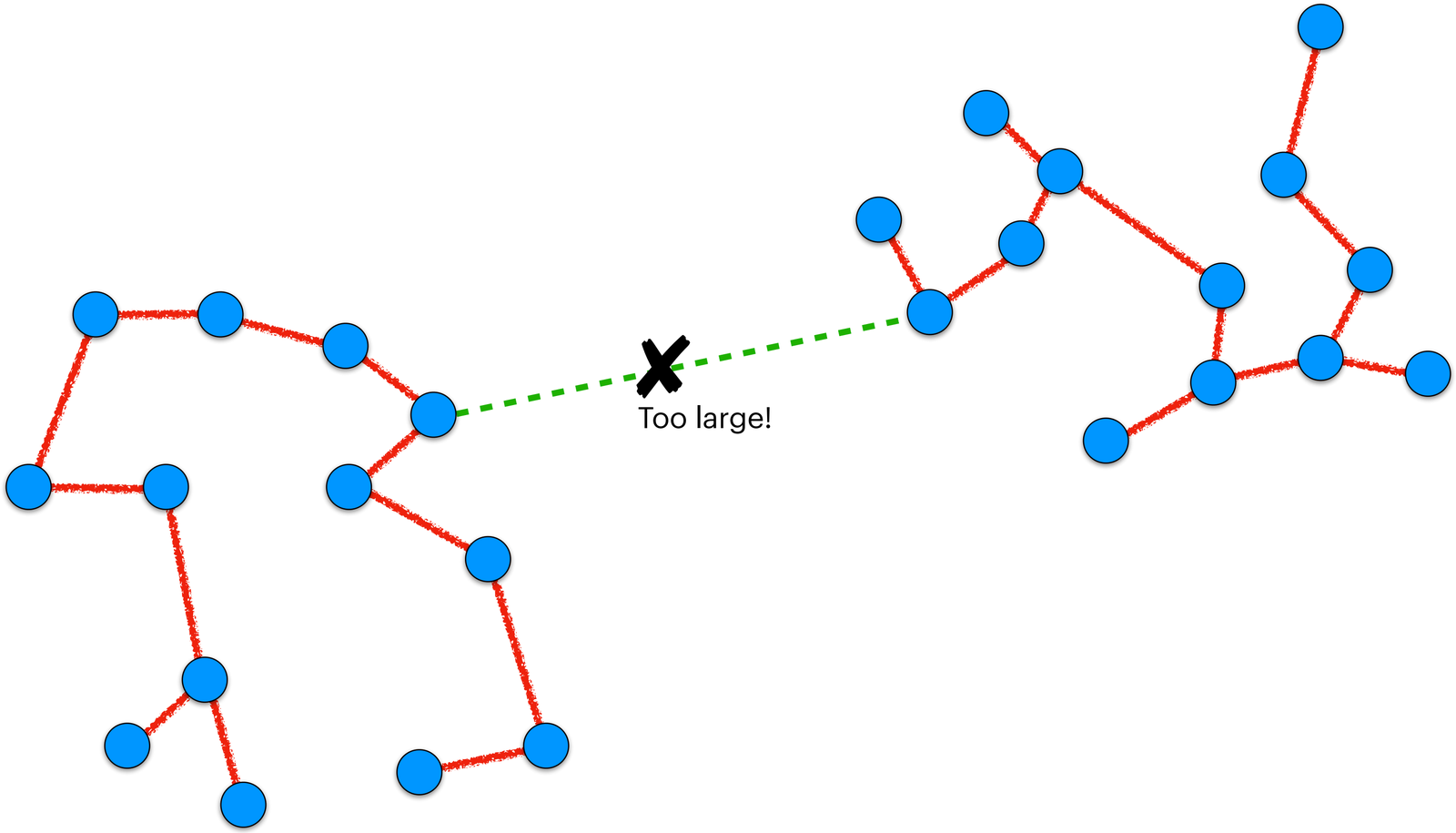


Algorithm

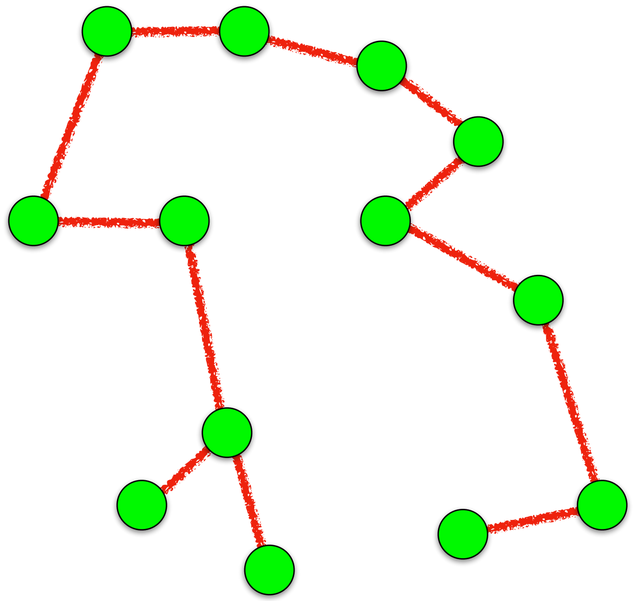


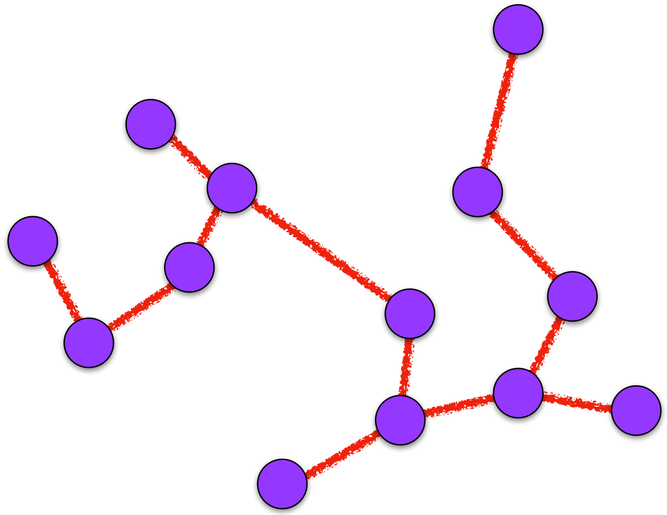


Algorithm

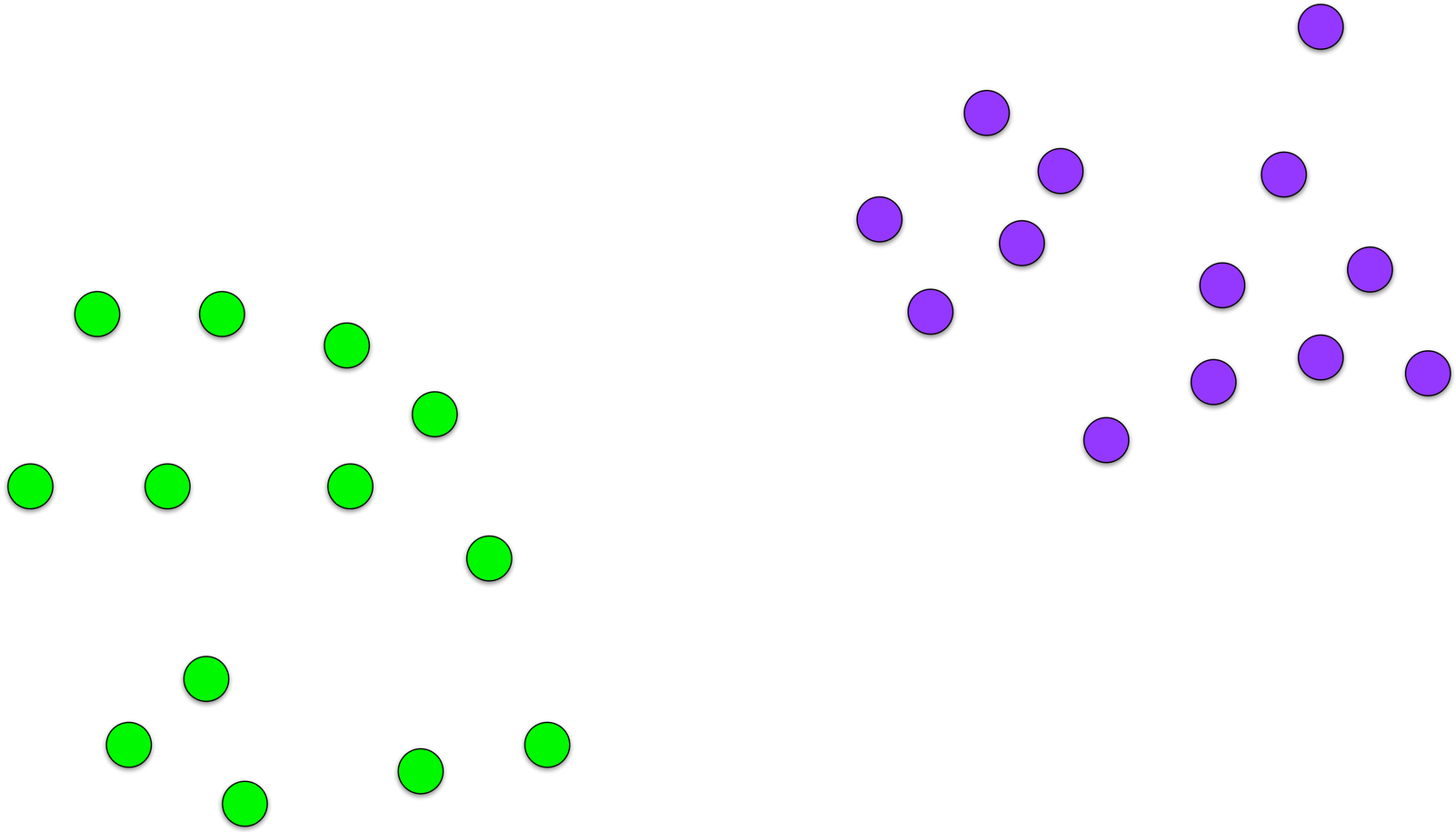


Algorithm

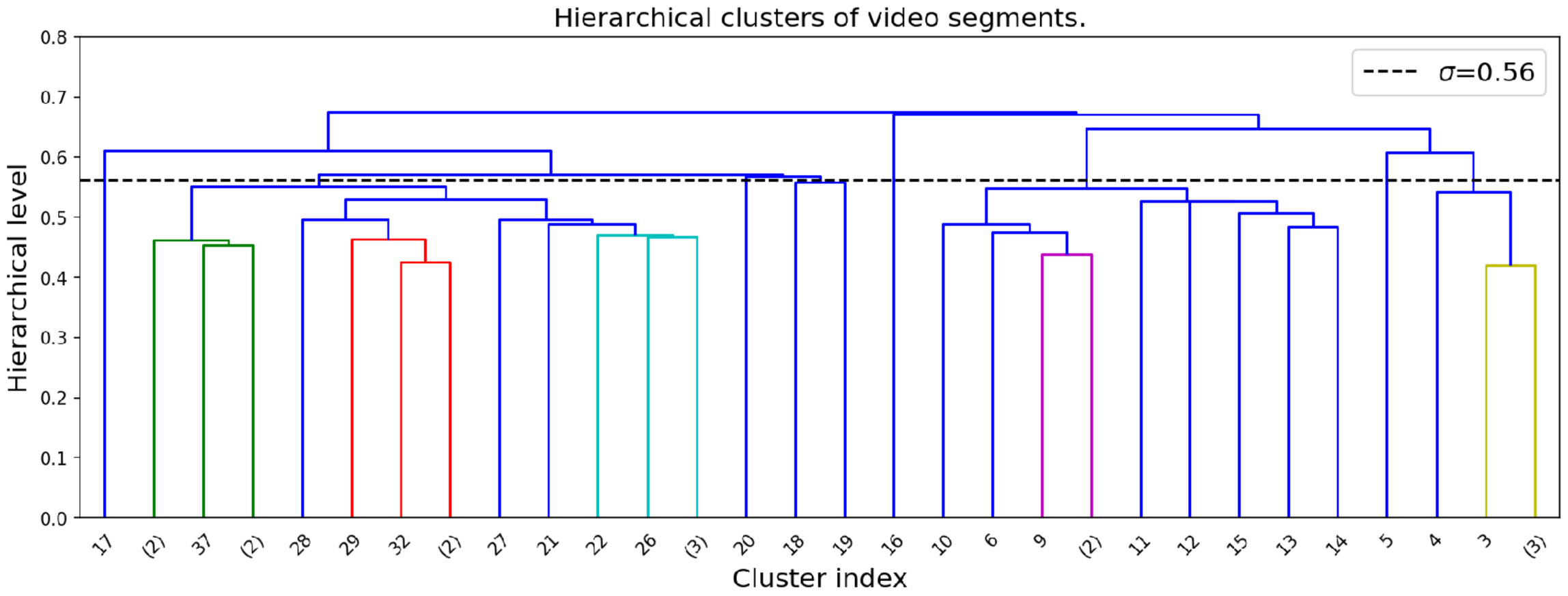




Algorithm



Algorithm: Dendogram



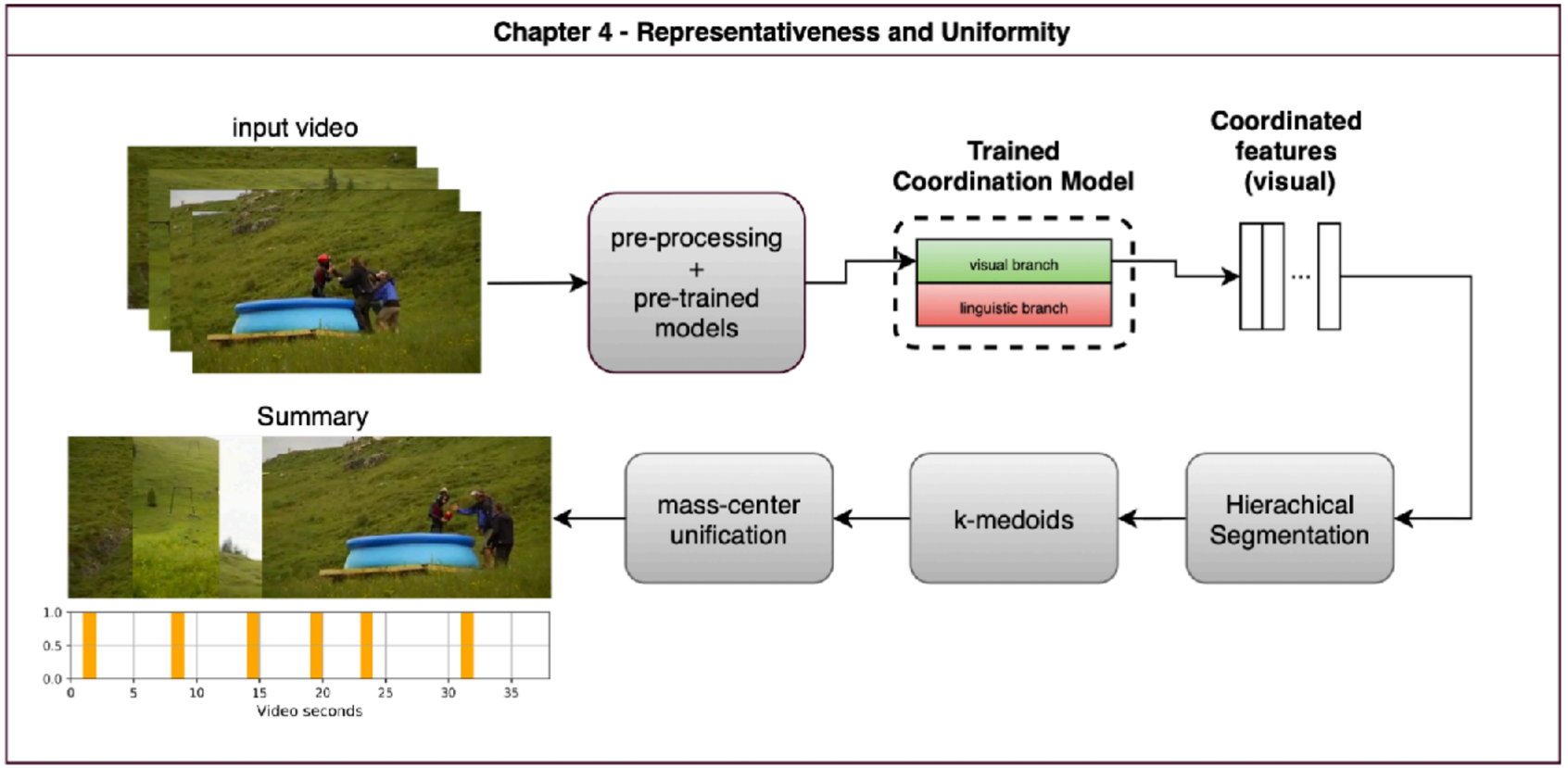
(Atencio et. al. , 2020) Query-based video summarization using machine learning and coordinated representations

Hierarchical Clustering

Single Linkage Clustering with Scikit-learn

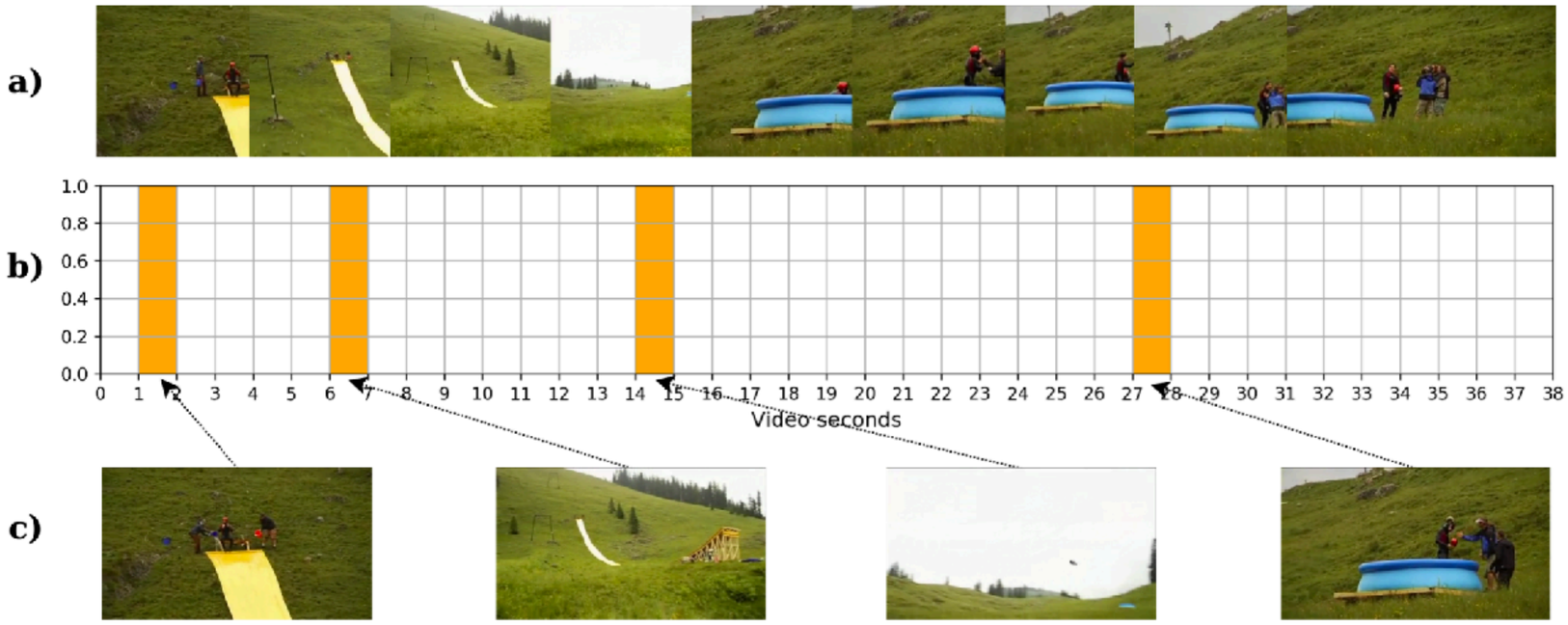
4. Study Case

Video Summarization

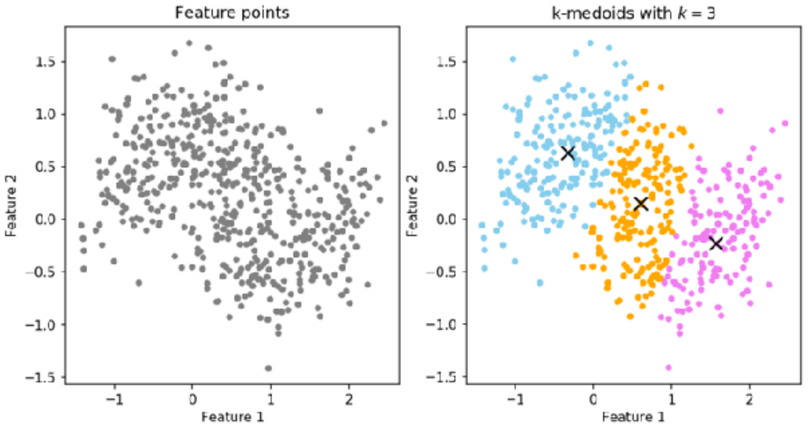


(Atencio et. al. , 2020) Query-based video summarization using machine learning and coordinated representations

Video Summarization



$$F(S) = \sum_{x \epsilon X} min_{s \epsilon S} \|x - S\|_2^2$$

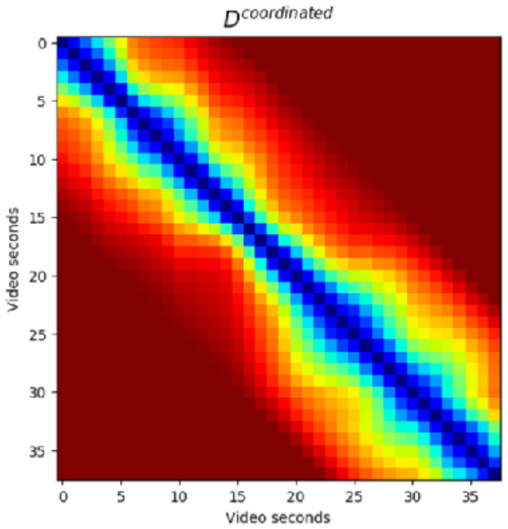


(Atencio et. al. , 2020) Query-based video summarization using machine learning and coordinated representations

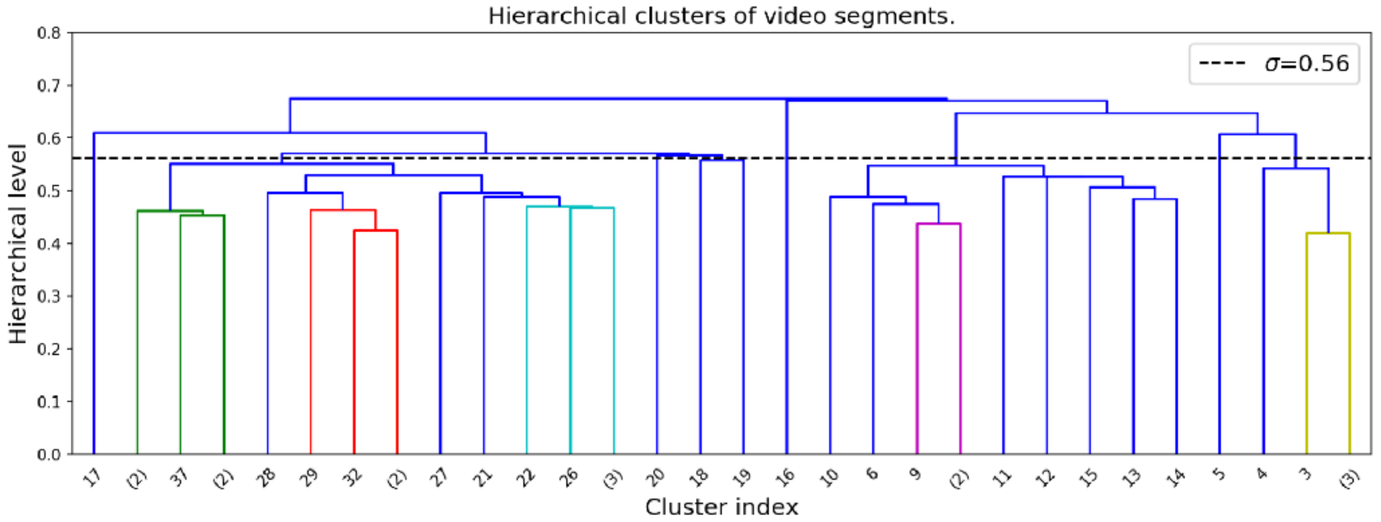
Video Summarization

$$w_{i,j}^{t} = \frac{1}{t} max(0, t - |i - j|)$$

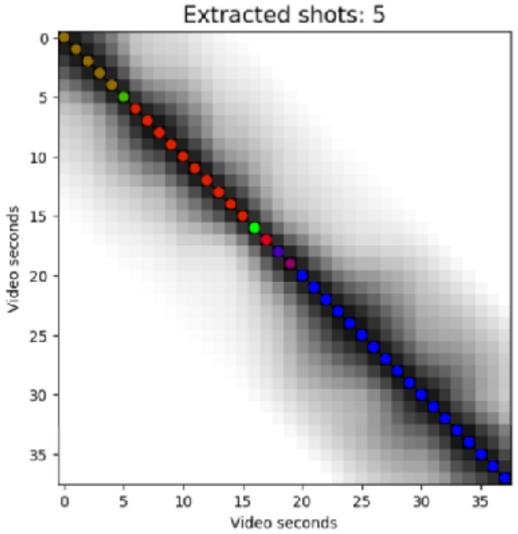
$$D_{i,j}^{coordinated}(C_i, C_j) = 1 - w_{i,j}^{t} e^{-\frac{1}{\Omega} cos(C_i, C_j)}$$



1) Disimilarity matrix.



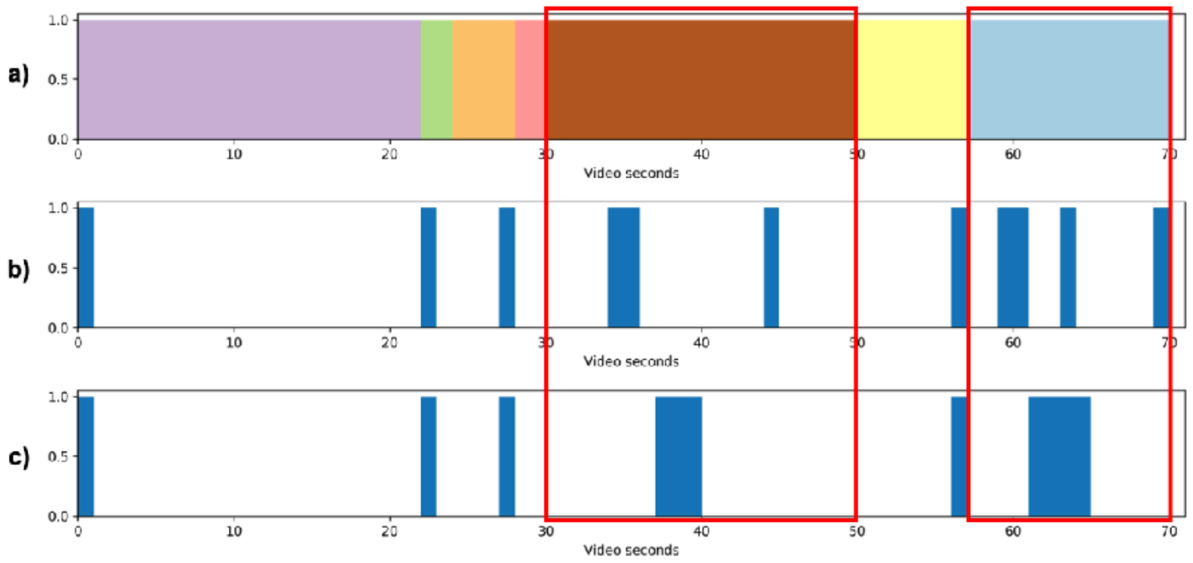
2) Hierarchical clustering and threshold σ.



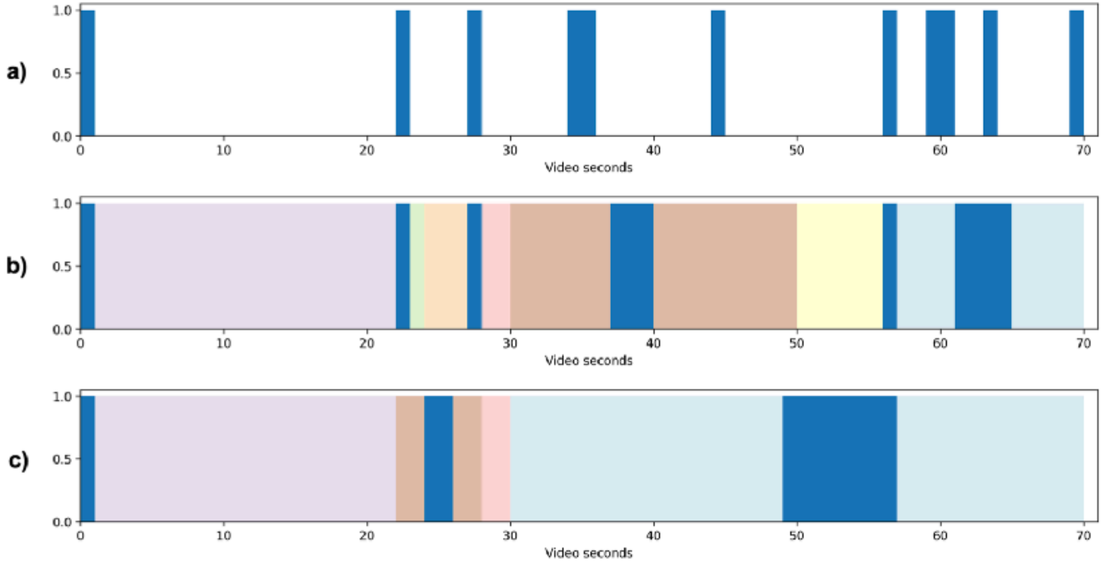
3) Extracted shots.

(Atencio et. al. , 2020) Query-based video summarization using machine learning and coordinated representations

Video Summarization



(a) Extracted shots,. (b) k-medoids over coordinated space.
(c) Uniformity by mass-center.

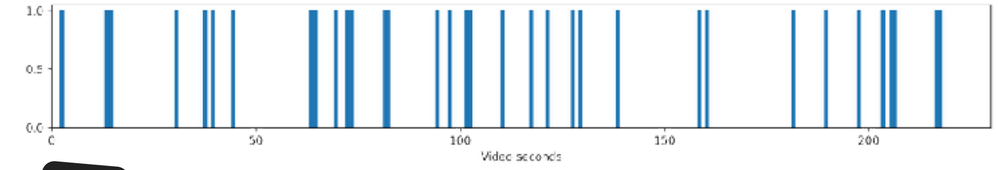


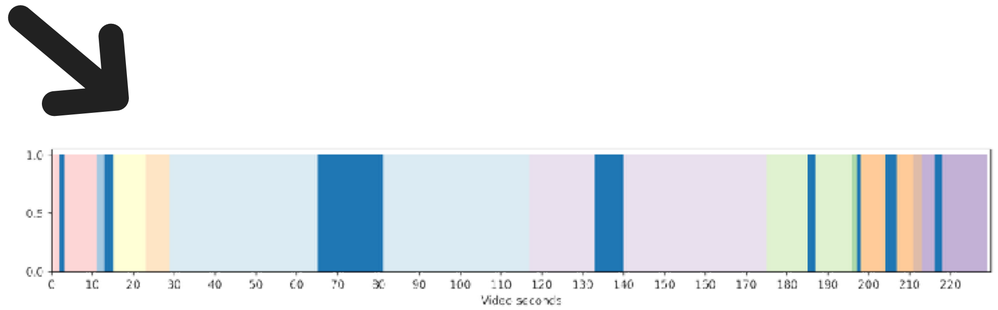
Sensitivity control using hierarchical clustering parameter σ.

(Atencio et. al. , 2020) Query-based video summarization using machine learning and coordinated representations

Video Summarization

Input Video





(Atencio et. al. , 2020) Query-based video summarization using machine learning and coordinated representations

Video Summarization

Representative Video Summarization (Live demonstration)

Conclusions

- Sequence summarization is a crucial task in computationally intensive task such as video analytics and text analytics.
- Extractive summarization is mainly unsupervised. On the contrary, abstractive summarization requires supervision, and therefore, human generated summaries.
- Hierarchical clustering allows to set sensitivity in extractive summarization tasks.