

USP — 21/11/22

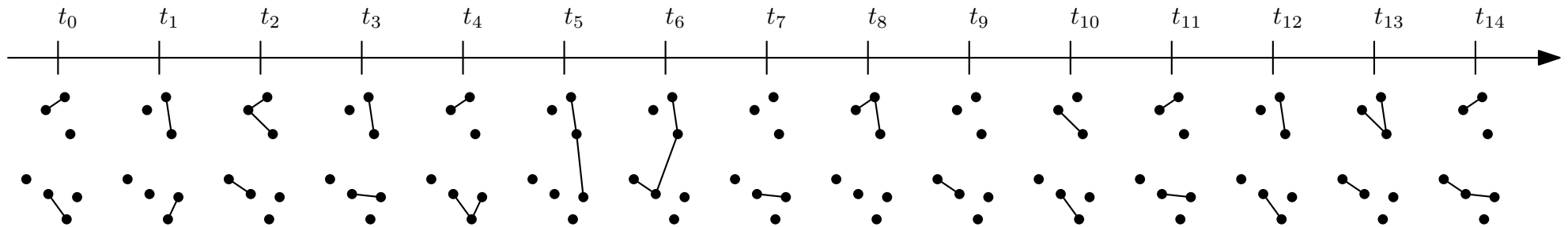
Grafos temporais e homologia persistente

Visualização e sugestão de resolução

Cláudio Linhares, Jean Ponciano and Raphaël Tinarrage

<https://raphaeltinarrage.github.io>

Um grafo temporal consiste nos dados de um grafo G e uma coleção de pares (v, t) , onde v é uma aresta de G e t um número real em $[0, T]$.

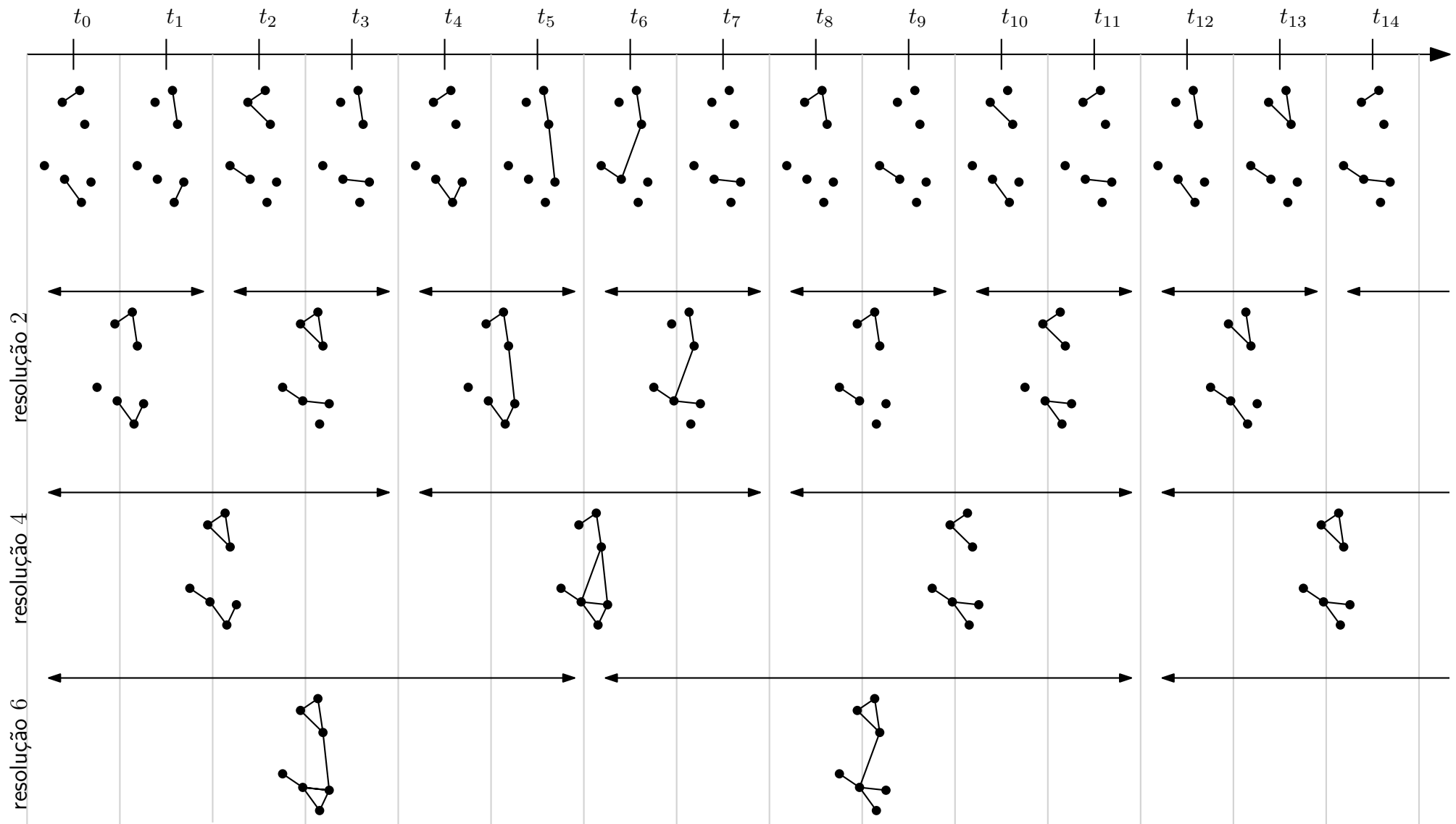


Esse formalismo está na base de muitos modelos de fenômenos dinâmicos (redes de comunicação, difusão de informação, mecanismos biológicos, ...).

Resolução

3/21 (1/2)

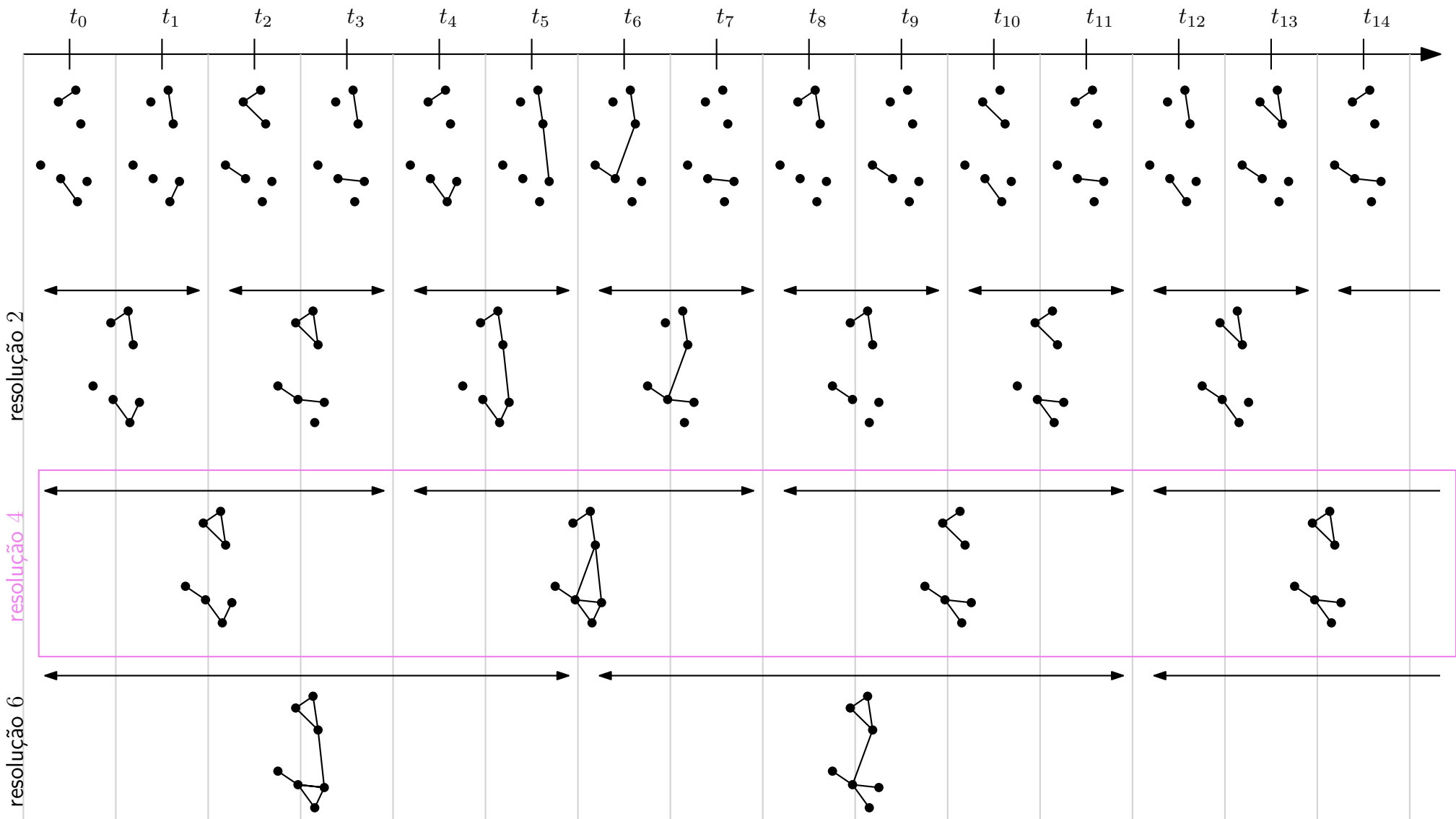
Dada uma **resolução** r , cortamos o intervalo $[0, T]$ em $n = \lceil T/r \rceil$ subintervalos $[kr, (k+1)r]$, e construímos grafos $\{G_k | k = 0, \dots, n-1\}$.



Resolução

3/21 (2/2)

Dada uma **resolução** r , cortamos o intervalo $[0, T]$ em $n = \lceil T/r \rceil$ subintervalos $[kr, (k+1)r]$, e construímos grafos $\{G_k | k = 0, \dots, n-1\}$.



Problema: Resoluções diferentes destacam comportamentos diferentes. Como escolher?

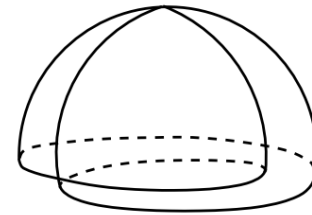
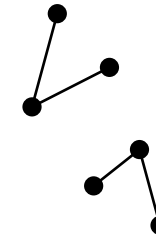
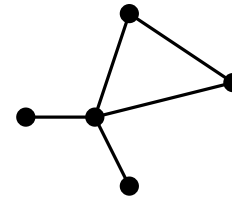
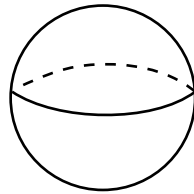
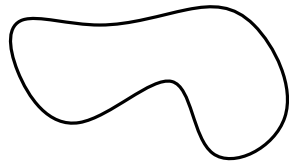
I - Homologia e homologia persistente

II - Código de barras colorido

III - Sugestão de resolução

IV - Exemplos

Seja X um espaço topológico.



subconjuntos de \mathbb{R}^n

grafos

variedade abstrata

Ele é associado a uma sequência de números inteiros, chamados de **números de Betti**. Esses números têm uma interpretação em termos de topologia.

$$\beta_0(X), \beta_1(X), \beta_2(X), \beta_3(X), \dots$$

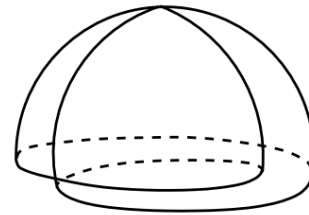
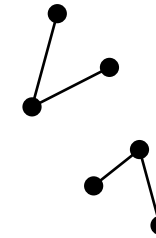
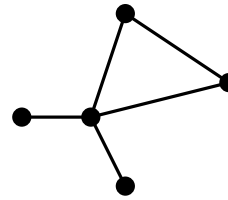
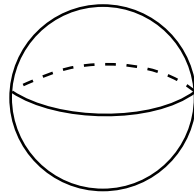
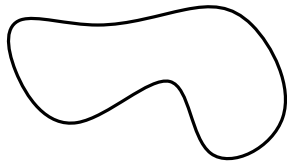
número de componentes
conexos

número buracos

número de cavidades

número de buracos de
dimensão 3

Seja X um espaço topológico.



subconjuntos de \mathbb{R}^n

grafos

variedade abstrata

$\beta_0(X)$	1	1	1	2	1
$\beta_1(X)$	1	0	1	0	1
$\beta_2(X)$	0	1	0	0	1

Ele é associado a uma sequência de números inteiros, chamados de **números de Betti**. Esses números têm uma interpretação em termos de topologia.

$$\beta_0(X), \beta_1(X), \beta_2(X), \beta_3(X), \dots$$

número de componentes
conexos

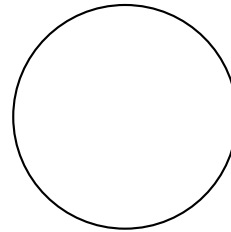
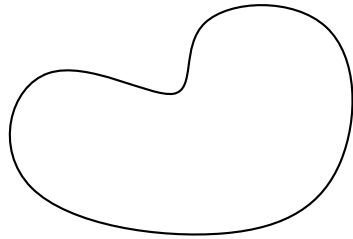
número buracos

número de cavidades

número de buracos de
dimensão 3

Em topologia, diz-se que dois espaços são equivalentes (homotopicamente equivalentes) se um puder ser **deformado** no outro.

Então, os números Betti são **invariantes topológicos**, ou seja, dois espaços equivalentes terão os mesmos números Betti



$$\beta_0(X) = \beta_0(Y) = 1$$

$$\beta_1(X) = \beta_1(Y) = 1$$

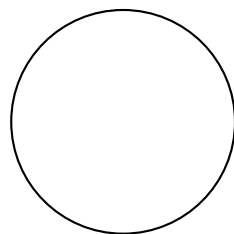
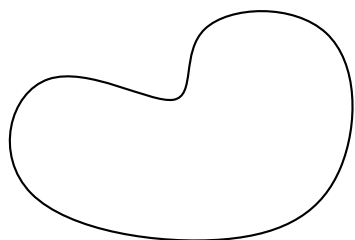
$$\beta_2(X) = \beta_2(Y) = 0$$

Invariância topológica

6/21 (2/3)

Em topologia, diz-se que dois espaços são equivalentes (homotopicamente equivalentes) se um puder ser **deformado** no outro.

Então, os números Betti são **invariantes topológicos**, ou seja, dois espaços equivalentes terão os mesmos números Betti

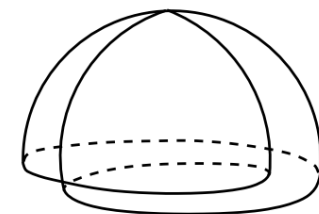
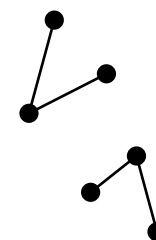
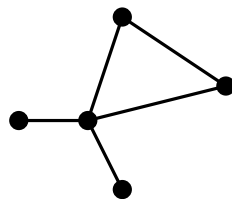
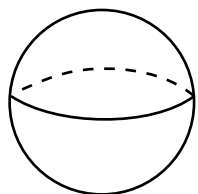
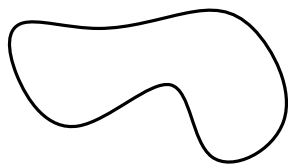


$$\beta_0(X) = \beta_0(Y) = 1$$

$$\beta_1(X) = \beta_1(Y) = 1$$

$$\beta_2(X) = \beta_2(Y) = 0$$

Logo, já sabemos quais espaços **não são equivalentes**.



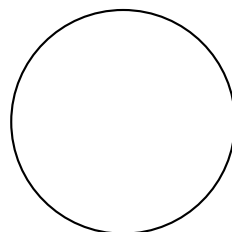
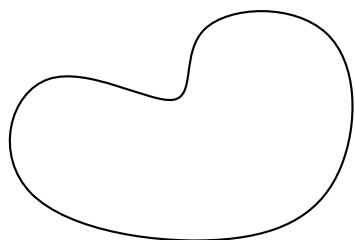
$\beta_0(X)$	1	1	1	2	1
$\beta_1(X)$	1	0	1	0	1
$\beta_2(X)$	0	1	0	0	1

Invariância topológica

6/21 (3/3)

Em topologia, diz-se que dois espaços são equivalentes (homotopicamente equivalentes) se um puder ser **deformado** no outro.

Então, os números Betti são **invariantes topológicos**, ou seja, dois espaços equivalentes terão os mesmos números Betti

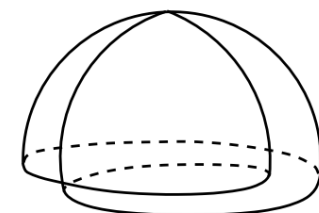
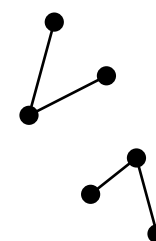
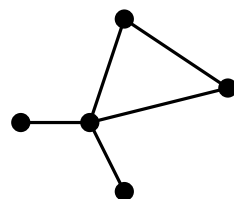
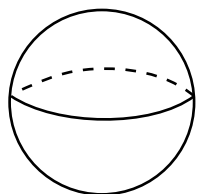
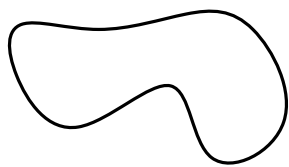


$$\beta_0(X) = \beta_0(Y) = 1$$

$$\beta_1(X) = \beta_1(Y) = 1$$

$$\beta_2(X) = \beta_2(Y) = 0$$

Logo, já sabemos quais espaços **não são equivalentes**.



$\beta_0(X)$

1

1

1

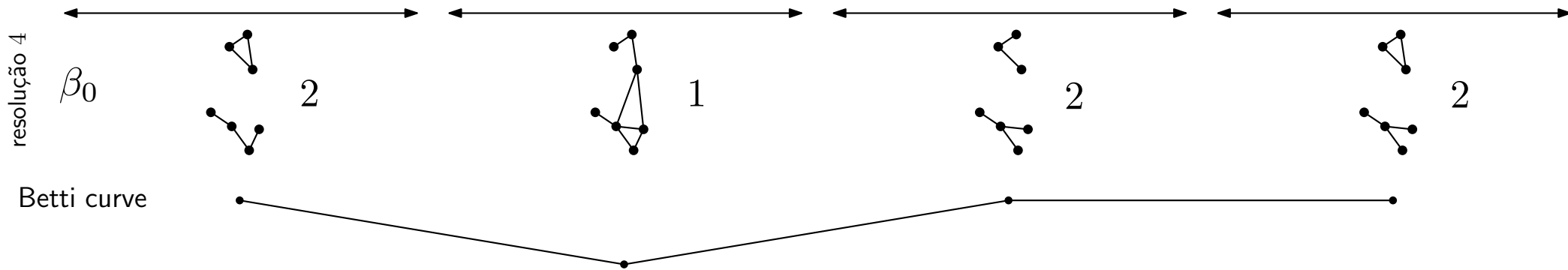
2

1

Em nosso trabalho, usamos apenas β_0 , o número de componentes conexos / grupos / clusters.

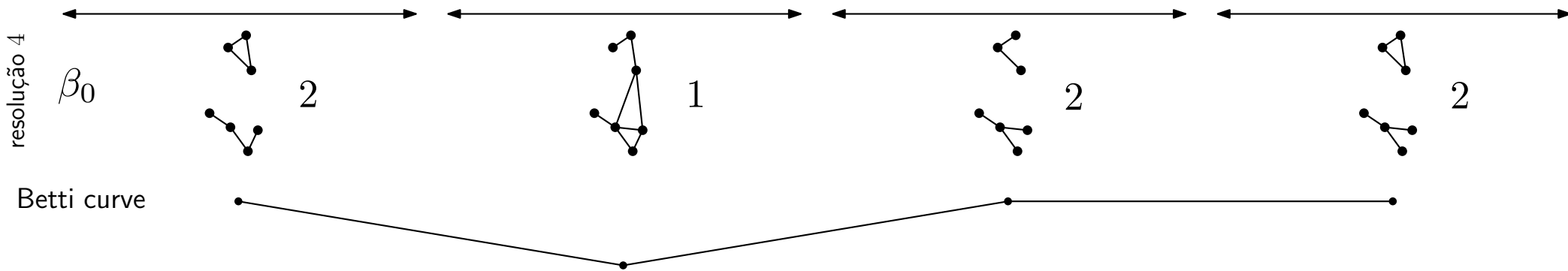
(H_0-) Homologia persistente zig-zag 7/21 (1/5)

Aqui, temos vários espaços topológicos. Desejamos entender a **evolução** dos clusters.

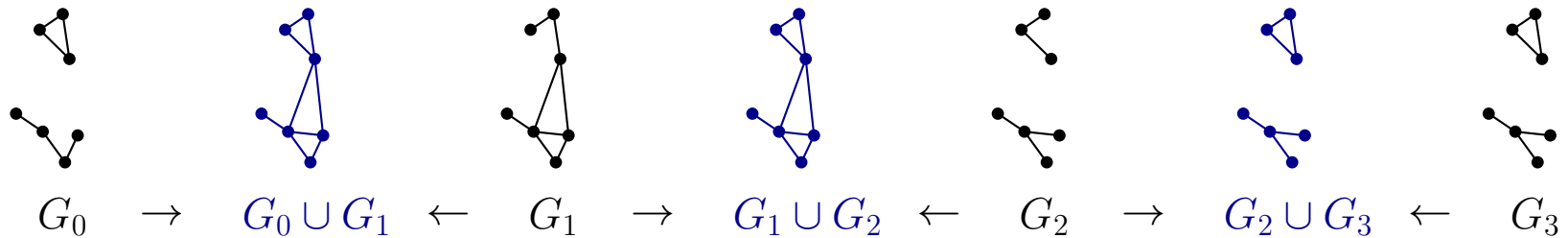


(H_0-) Homologia persistente zig-zag 7/21 (2/5)

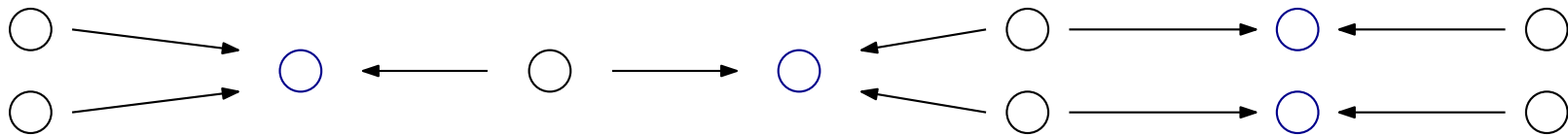
Aqui, temos vários espaços topológicos. Desejamos entender a **evolução** dos clusters.



Podemos construir a filtração zig-zag associada.

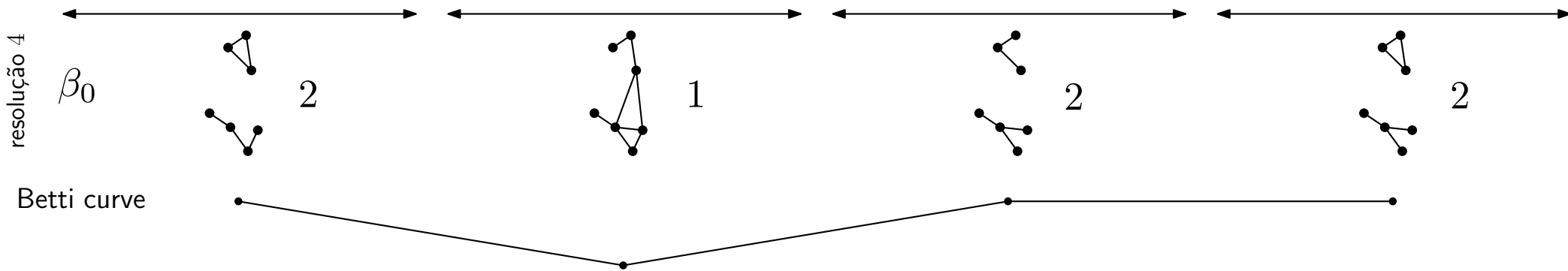


No *merge tree*, podemos acompanhar a evolução dos componentes conexos.

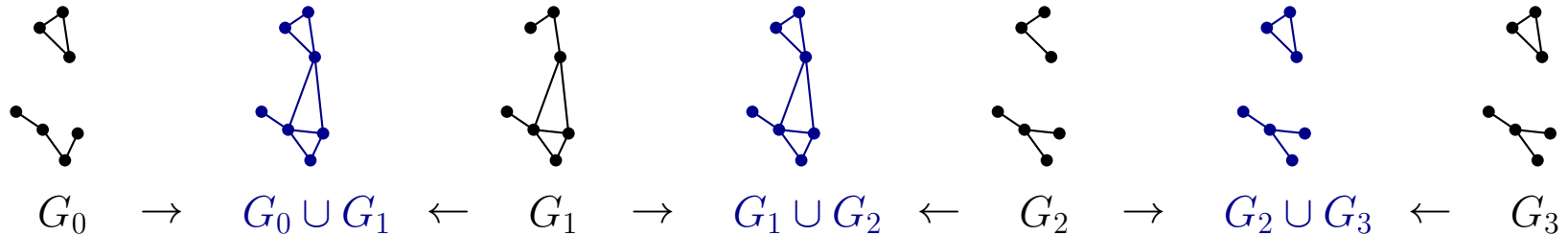


(H_0-) Homologia persistente zig-zag 7/21 (3/5)

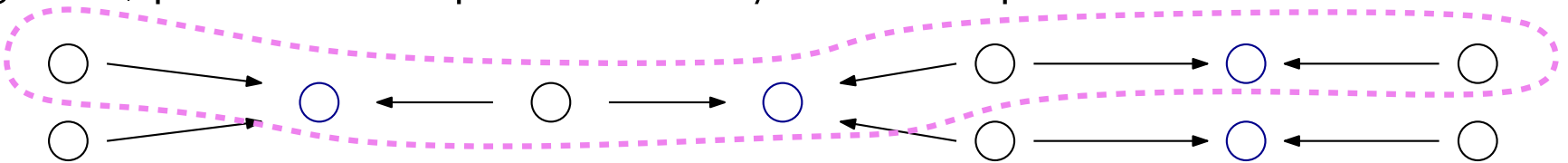
Aqui, temos vários espaços topológicos. Desejamos entender a **evolução** dos clusters.



Podemos construir a filtração zig-zag associada.

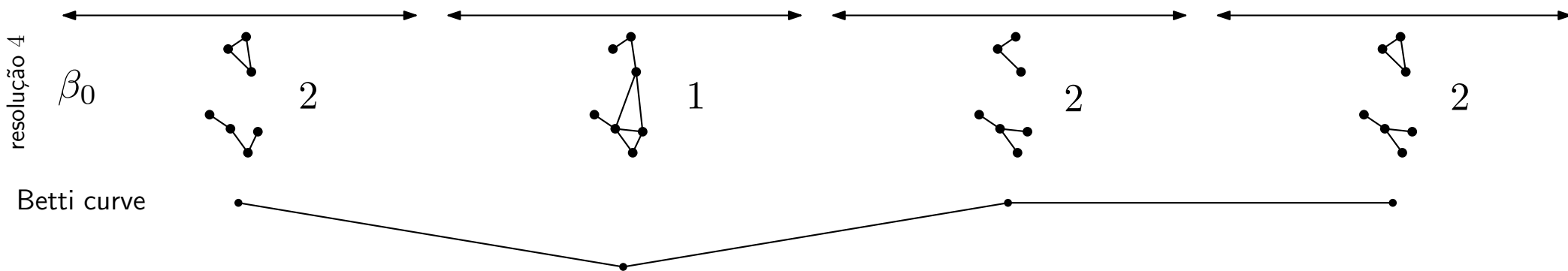


No *merge tree*, podemos acompanhar a evolução dos componentes conexos.

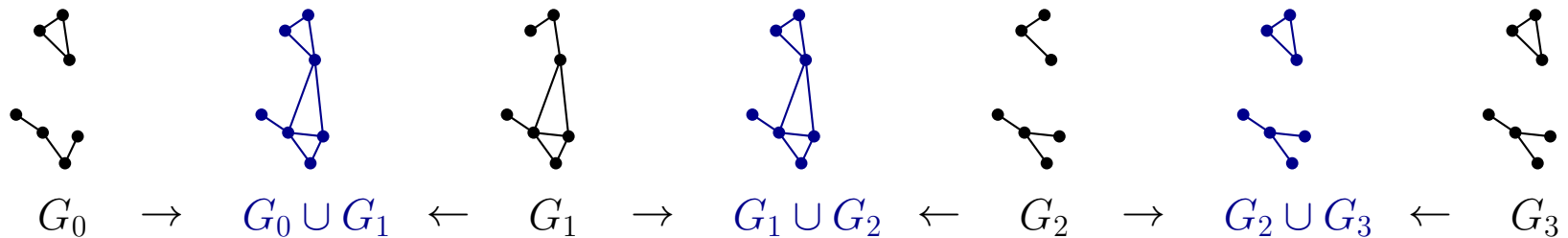


(H_0-) Homologia persistente zig-zag 7/21 (4/5)

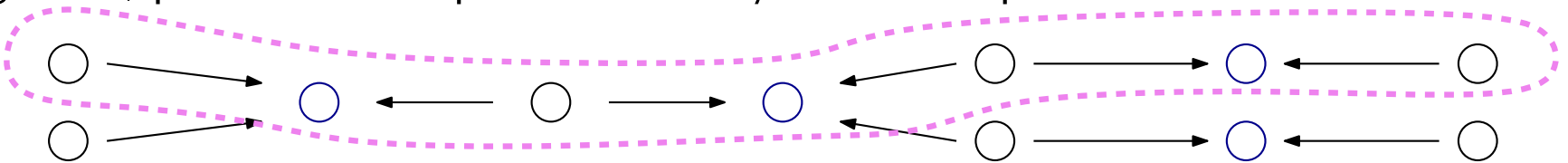
Aqui, temos vários espaços topológicos. Desejamos entender a **evolução** dos clusters.



Podemos construir a filtração zig-zag associada.



No *merge tree*, podemos acompanhar a evolução dos componentes conexos.



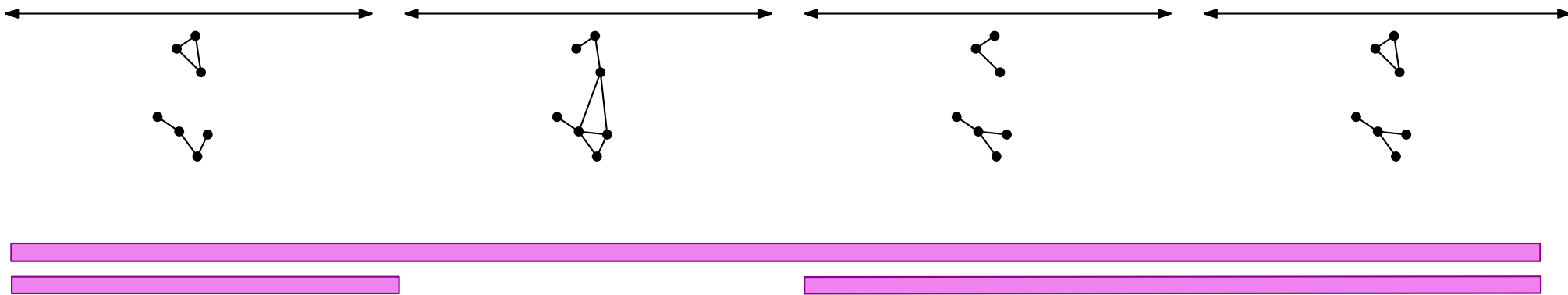
Isto permite definir o **código de barras**.



(H_0-) Homologia persistente zig-zag 7/21 (5/5)

Um código de barras é um conjunto de intervalos $\{[b_i, d_i] \mid i\}$. A construção dele é baseada na teoria da homologia persistente.

Num código de barras, vemos tanto os componentes conexos a cada momento, mas também a **persistência** deles.



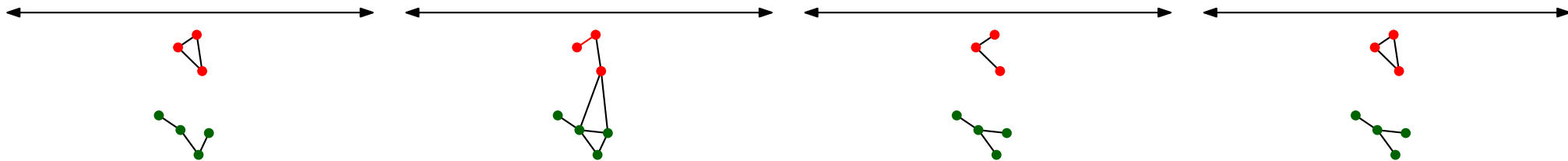
I - Homologia e homologia persistente

II - Código de barras colorido

III - Sugestão de resolução

IV - Exemplos

Nossa primeira contribuição: o código de barras colorido. Baseado em classes. Ele permite identificar a composição de cada um dos componentes conexos.



Código de barras não colorido

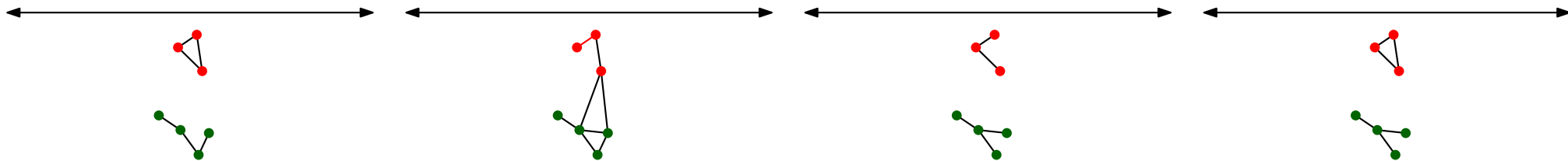


Código de barras colorido



A construção é baseada em [Computing zigzag persistence on graphs in near-linear time, Tamal Dey e Tao Hou].

Nossa primeira contribuição: o código de barras colorido. Baseado em classes. Ele permite identificar a composição de cada um dos componentes conexos.



Código de barras não colorido



Código de barras colorido



A construção é baseada em [Computing zigzag persistence on graphs in near-linear time, Tamal Dey e Tao Hou].

Observação: o código de barras colorido não é único.



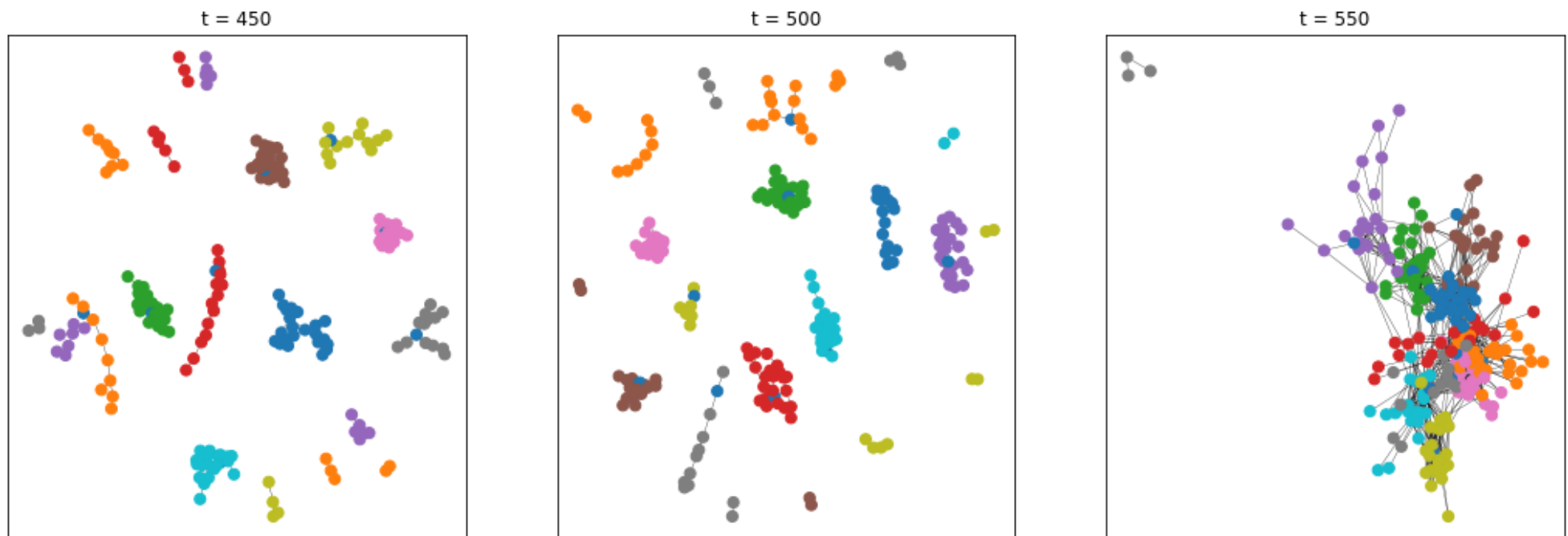
Exemplo: primary school dataset

10/21 (1/4)

Este conjunto de dados representa as conexões entre os alunos de uma escola primária durante uma semana.



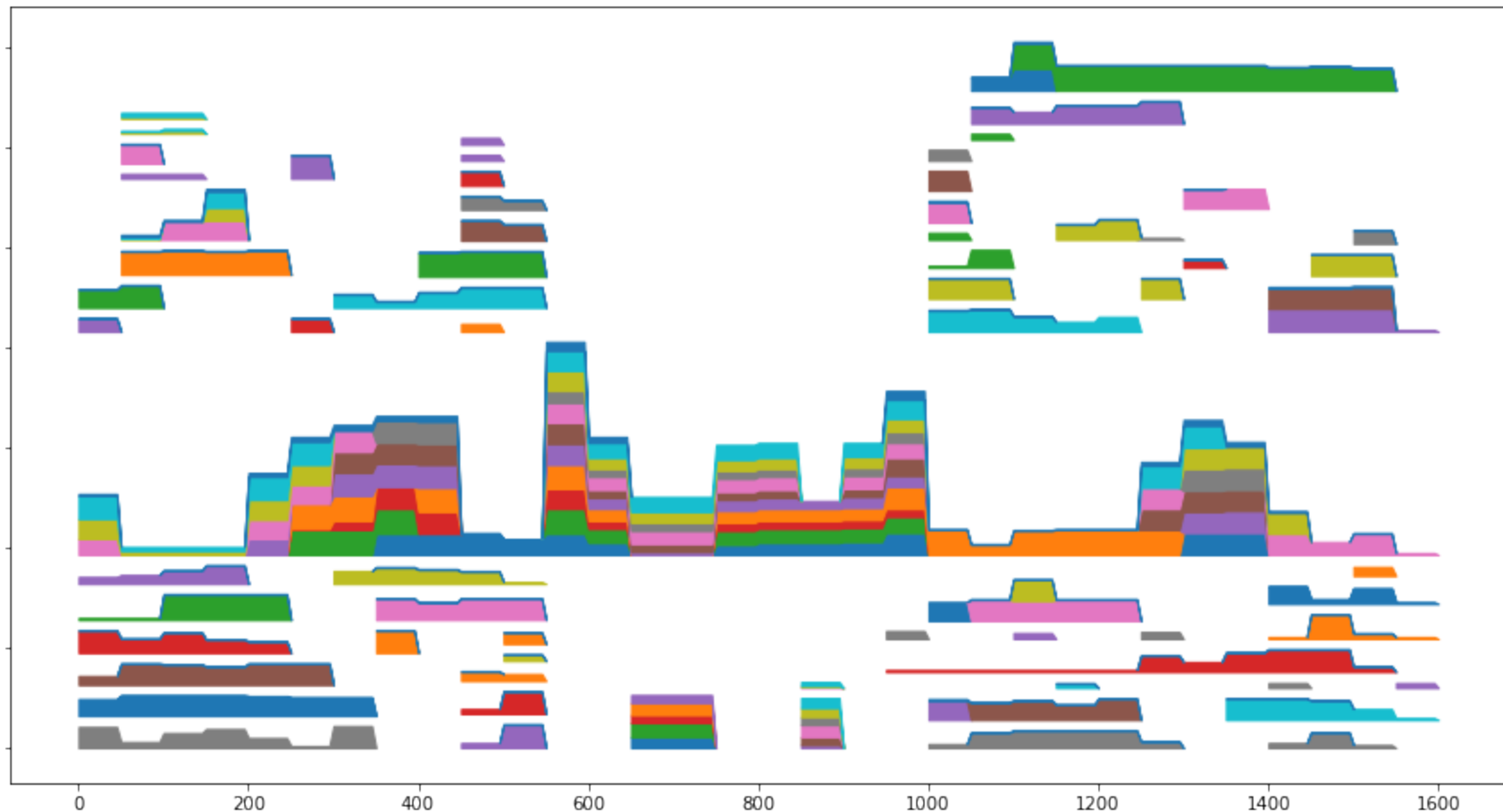
Com resolução de 50 ($\simeq 16$ minutos):



Exemplo: primary school dataset

10/21 (2/4)

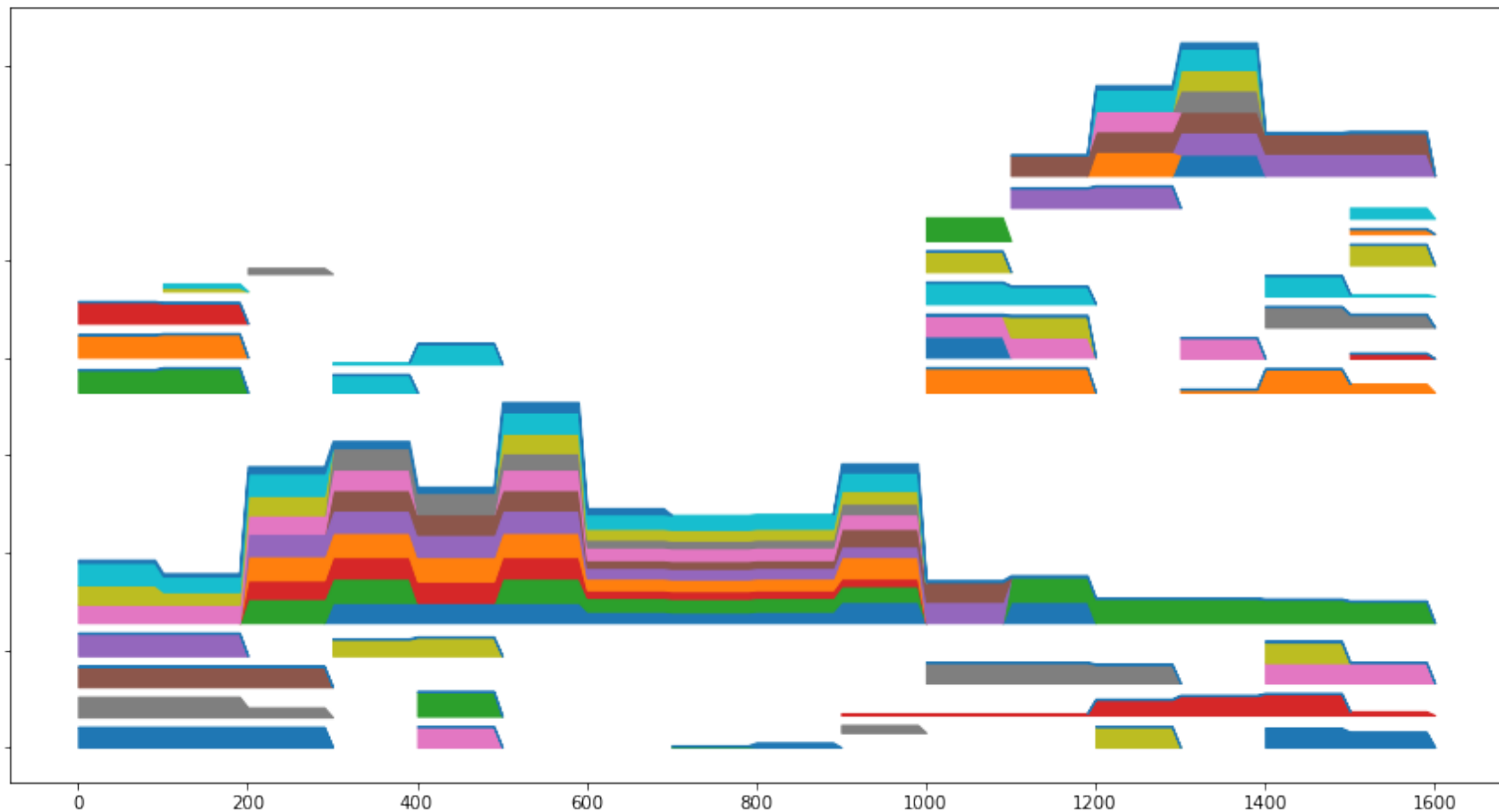
Este conjunto de dados representa as conexões entre os alunos de uma escola primária durante uma semana.



Exemplo: primary school dataset

10/21 (3/4)

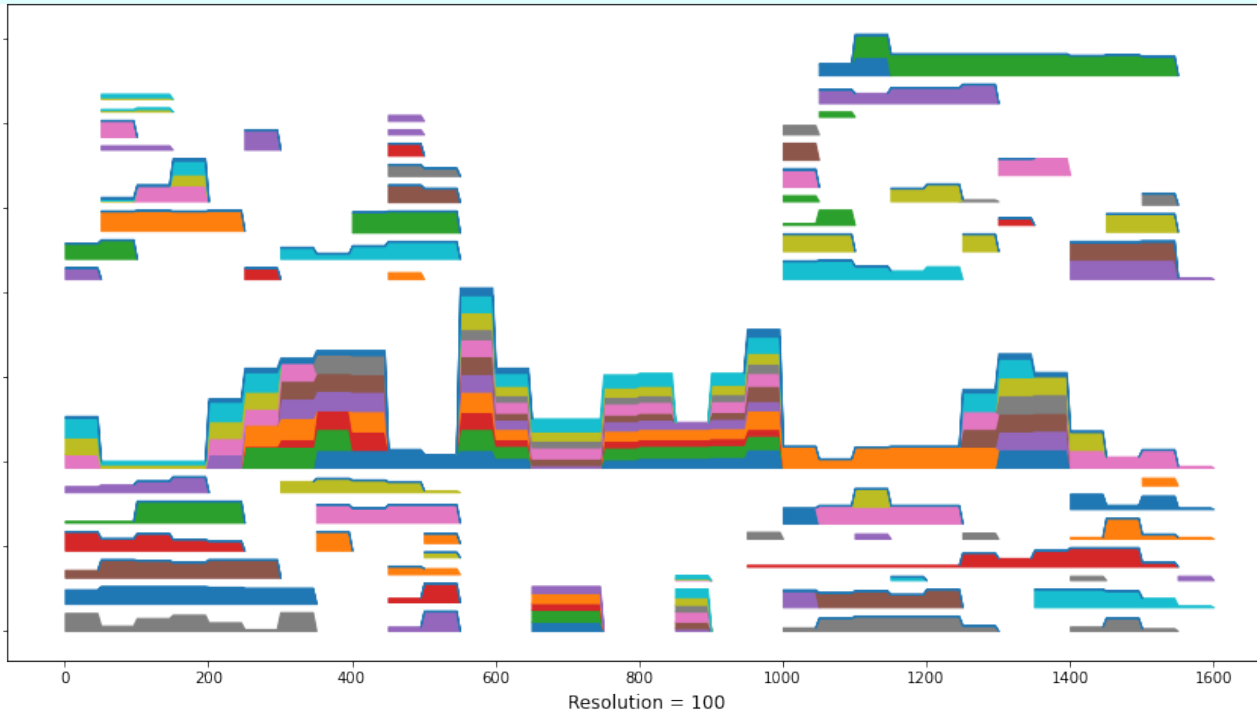
Este conjunto de dados representa as conexões entre os alunos de uma escola primária durante uma semana.



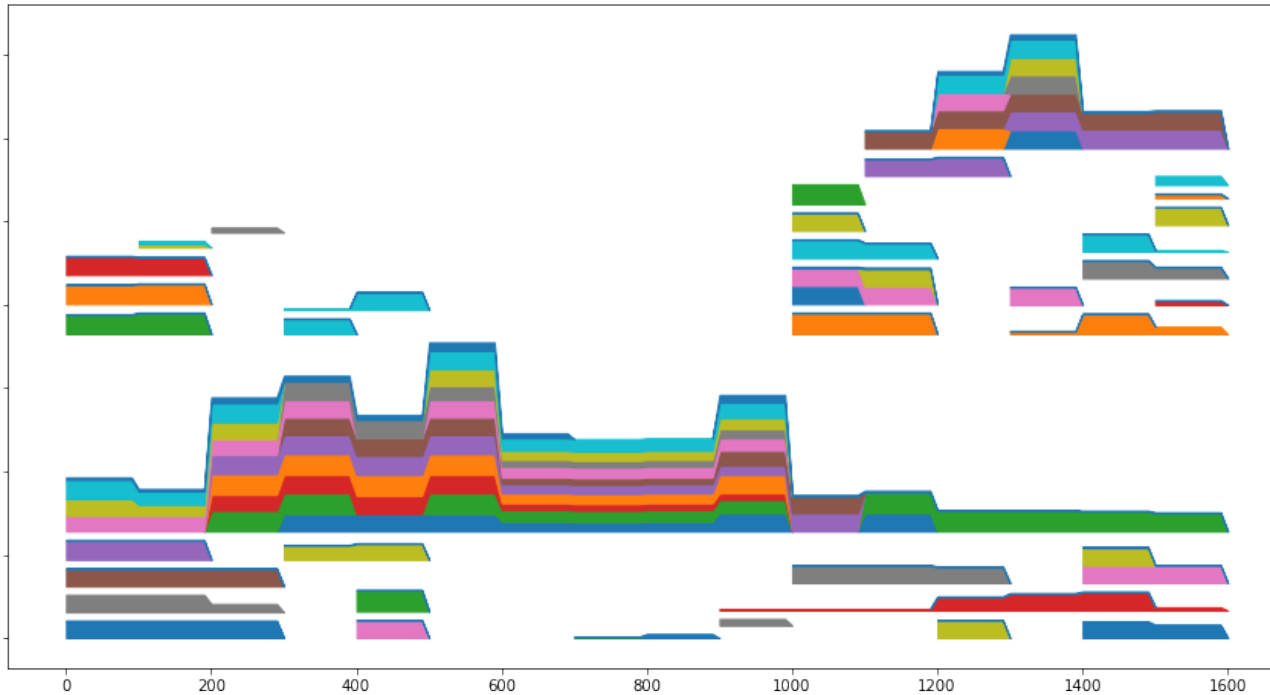
Exemplo: primary school dataset

10/21 (4/4)

Resolution = 50



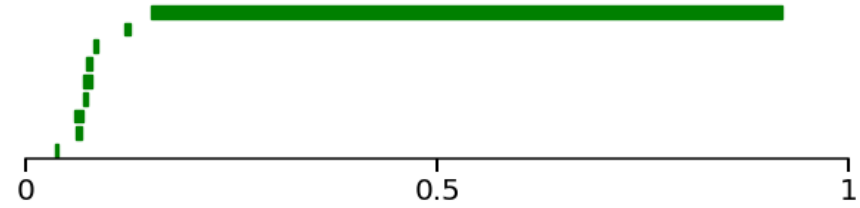
Resolution = 100



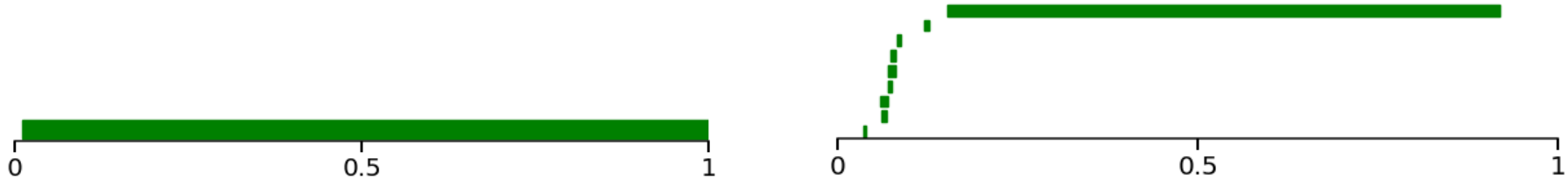
Distância bottleneck

11/21 (1/8)

Sejam dois códigos de barras P e Q .



Sejam dois códigos de barras P e Q .



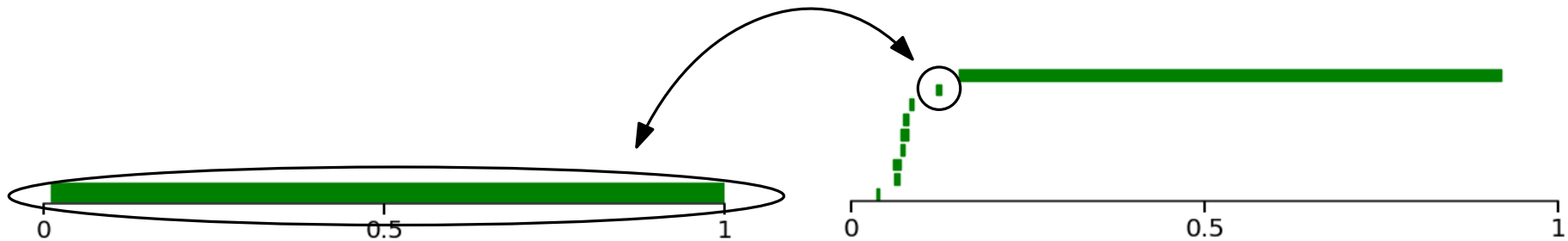
Uma **correspondência parcial** entre P e Q e um subconjunto $M \subset P \times Q$ tal como

- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M .

Se uma barra $p \in P$ (resp. $q \in Q$) não for correspondida por M , digamos que é correspondida com o conjunto unitário $\bar{p} = \left[\frac{p_1 + p_2}{2}, \frac{p_1 + p_2}{2} \right]$ (resp. $\bar{q} = \left[\frac{q_1 + q_2}{2}, \frac{q_1 + q_2}{2} \right]$).

Sejam dois códigos de barras P e Q .



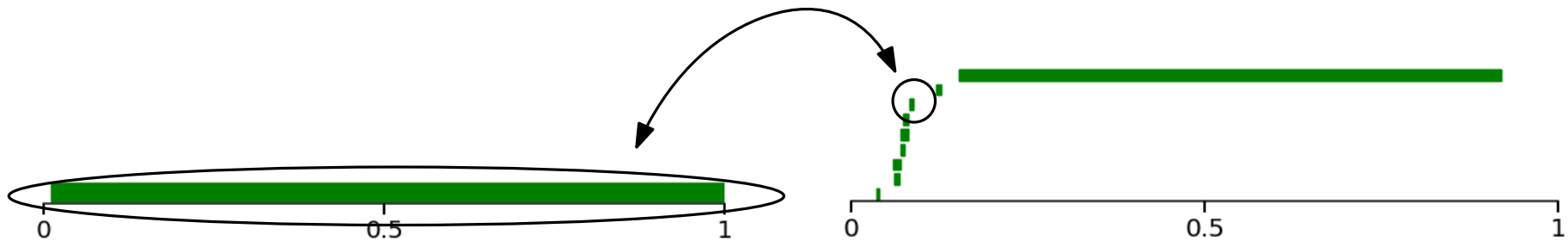
Uma **correspondência parcial** entre P e Q e um subconjunto $M \subset P \times Q$ tal como

- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M .

Se uma barra $p \in P$ (resp. $q \in Q$) não for correspondida por M , digamos que é correspondida com o conjunto unitário $\bar{p} = \left[\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2} \right]$ (resp. $\bar{q} = \left[\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2} \right]$).

Sejam dois códigos de barras P e Q .



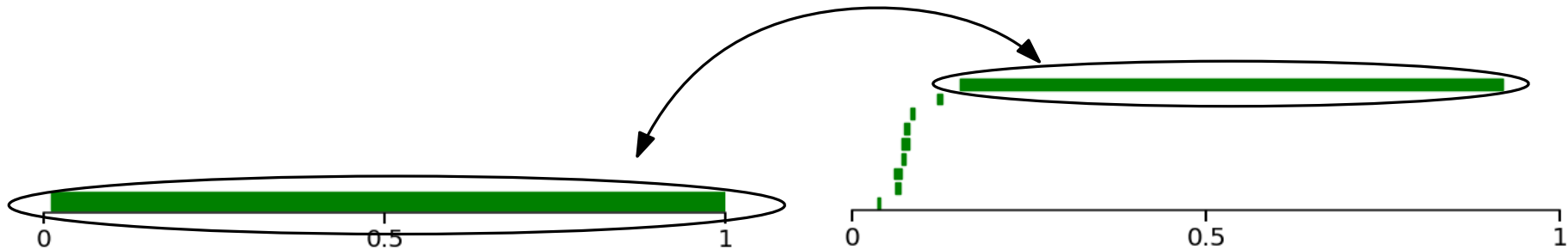
Uma **correspondência parcial** entre P e Q e um subconjunto $M \subset P \times Q$ tal como

- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M .

Se uma barra $p \in P$ (resp. $q \in Q$) não for correspondida por M , digamos que é correspondida com o conjunto unitário $\bar{p} = \left[\frac{p_1 + p_2}{2}, \frac{p_1 + p_2}{2} \right]$ (resp. $\bar{q} = \left[\frac{q_1 + q_2}{2}, \frac{q_1 + q_2}{2} \right]$).

Sejam dois códigos de barras P e Q .



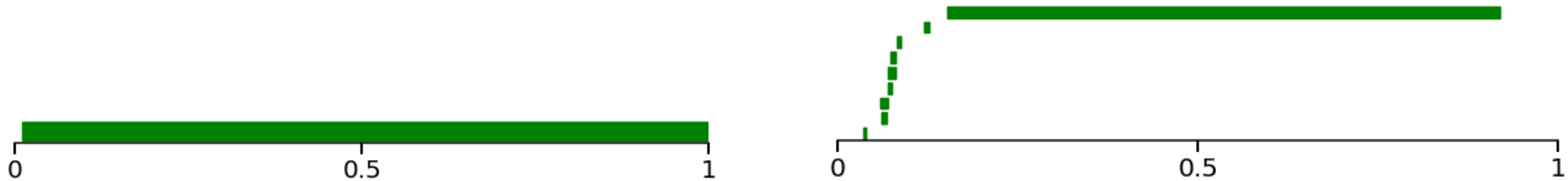
Uma **correspondência parcial** entre P e Q e um subconjunto $M \subset P \times Q$ tal como

- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M .

Se uma barra $p \in P$ (resp. $q \in Q$) não for correspondida por M , digamos que é correspondida com o conjunto unitário $\bar{p} = \left[\frac{p_1 + p_2}{2}, \frac{p_1 + p_2}{2} \right]$ (resp. $\bar{q} = \left[\frac{q_1 + q_2}{2}, \frac{q_1 + q_2}{2} \right]$).

Sejam dois códigos de barras P e Q .



Uma **correspondência parcial** entre P e Q e um subconjunto $M \subset P \times Q$ tal como

- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

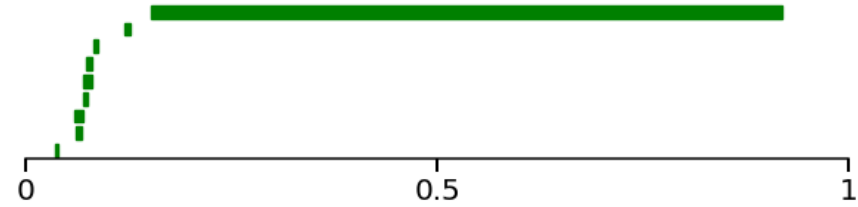
As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M .

Se uma barra $p \in P$ (resp. $q \in Q$) não for correspondida por M , digamos que é correspondida com o conjunto unitário $\bar{p} = \left[\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2} \right]$ (resp. $\bar{q} = \left[\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2} \right]$).

O **custo** de um par (p, q) (resp. (p, \bar{p}) , resp. (\bar{q}, q)) é a norma sup $\|p - q\|_\infty = \sup\{|p_1 - q_1|, |p_2 - q_2|\}$ (resp. $\|p - \bar{p}\|_\infty$, resp. $\|\bar{q} - q\|_\infty$).

O **custo** da correspondência M , referido como $\text{cost}(M)$, é o supremo dos custos das pares.

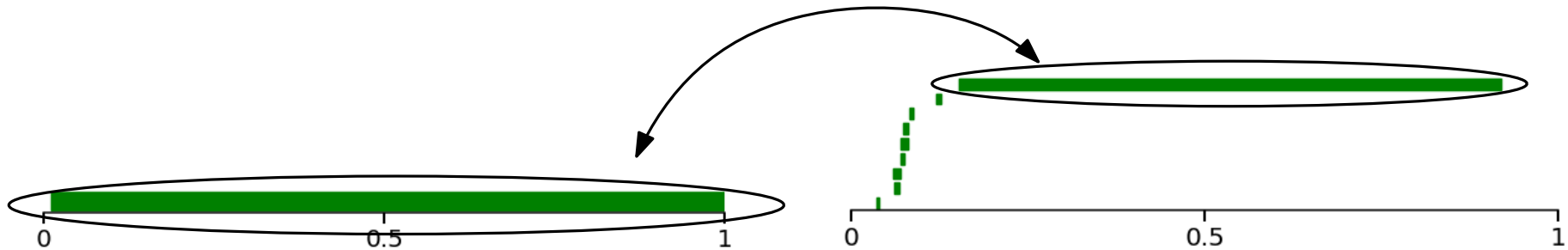
Sejam dois códigos de barras P e Q .



Definição : A **distância bottleneck** entre P e Q é definida como o menor custo de qualquer correspondência:

$$d_b(P, Q) = \inf\{\text{cost}(M), M \text{ é uma correspondência entre } P \text{ e } Q\}.$$

Sejam dois códigos de barras P e Q .



Definição : A **distância bottleneck** entre P e Q é definida como o menor custo de qualquer correspondência:

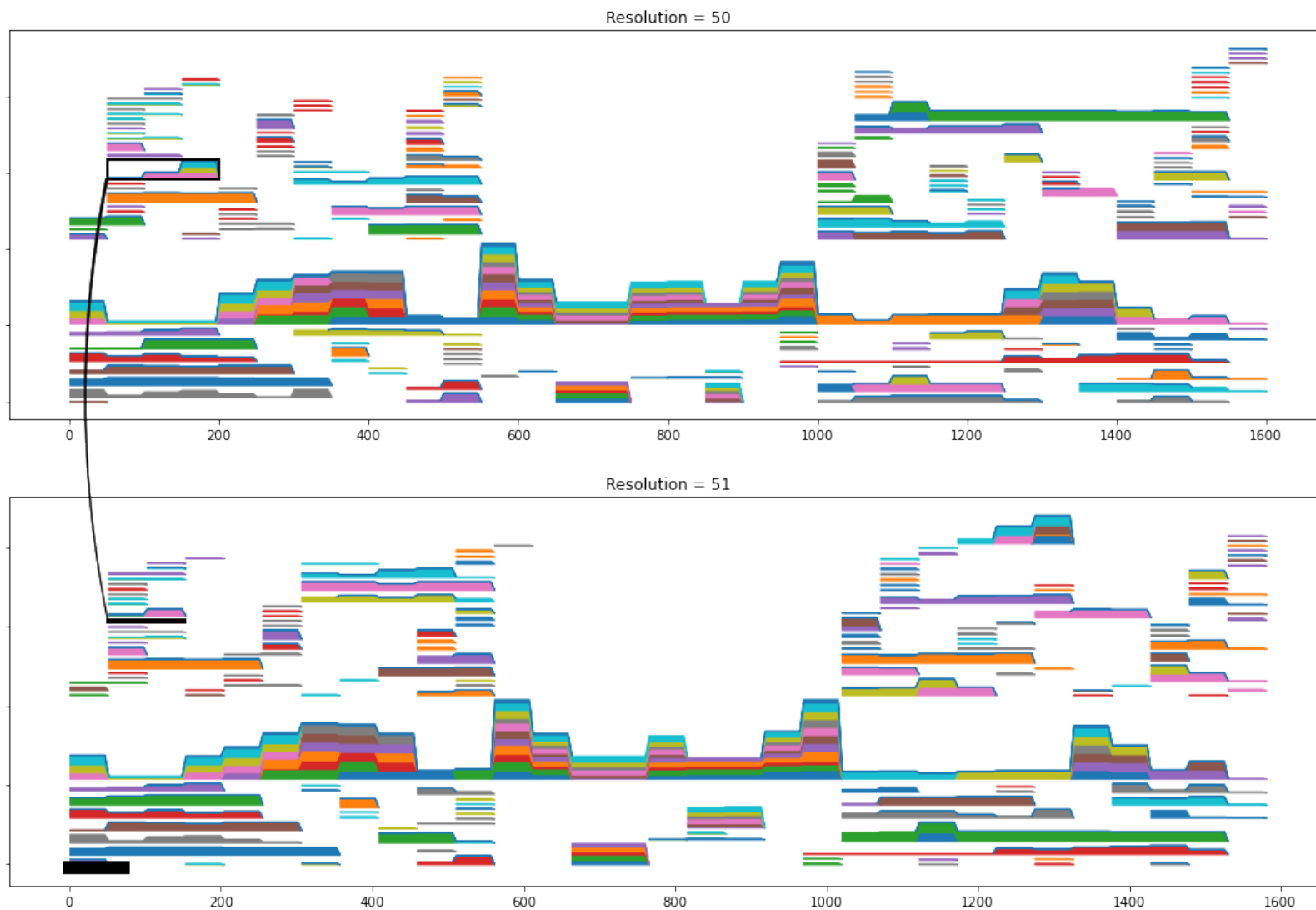
$$d_b(P, Q) = \inf\{\text{cost}(M), M \text{ é uma correspondência entre } P \text{ e } Q\}.$$

Mudanças locais e globais

12/21 (1/2)

Indo da resolução r_i à r_{i+1} , se distinguem dois valores de distância:

- se $d(r_i, r_{i+1}) \leq r_{i+1}$, então é uma **modificação local** do código de barras. Não é relevante.
- se $d(r_i, r_{i+1}) > r_{i+1}$, então é uma **mudança estrutural**. Isto significa que entramos em outro comportamento do grafo dinâmico.

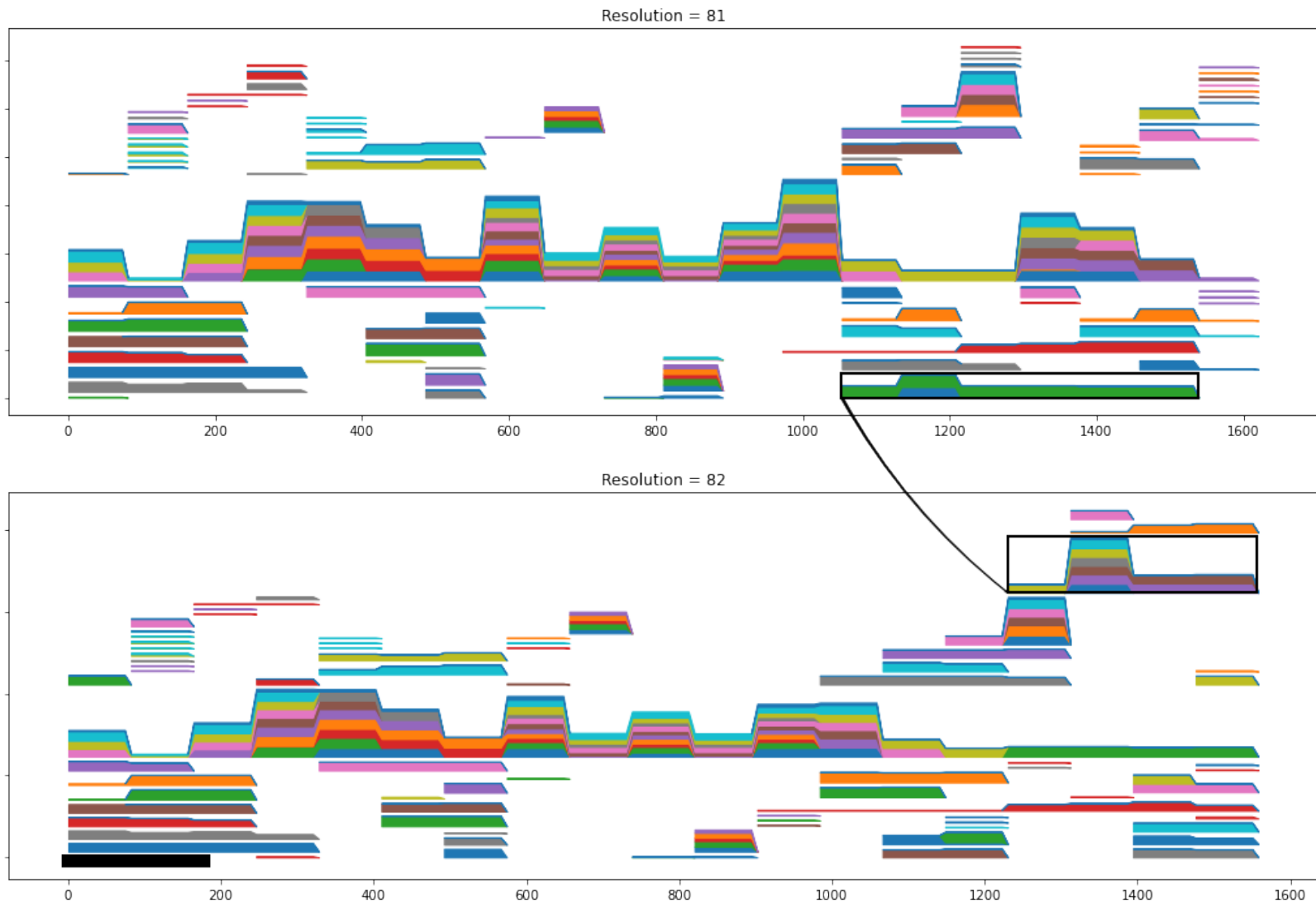


Mudanças locais e globais

12/21 (2/2)

Indo da resolução r_i à r_{i+1} , se distinguem dois valores de distância:

- se $d(r_i, r_{i+1}) \leq r_{i+1}$, então é uma **modificação local** do código de barras. Não é relevante.
- se $d(r_i, r_{i+1}) > r_{i+1}$, então é uma **mudança estrutural**. Isto significa que entramos em outro comportamento do grafo dinâmico.



I - Homologia e homologia persistente

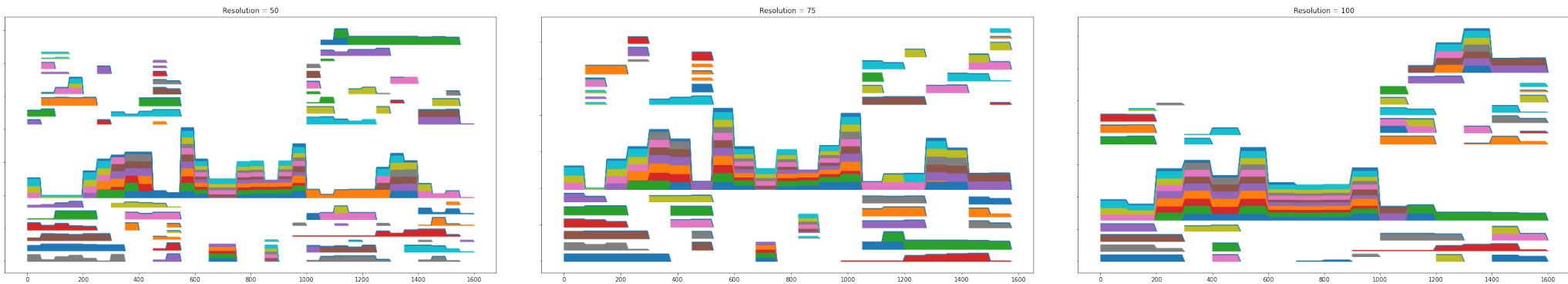
II - Código de barras decorado

III - Sugestão de resolução

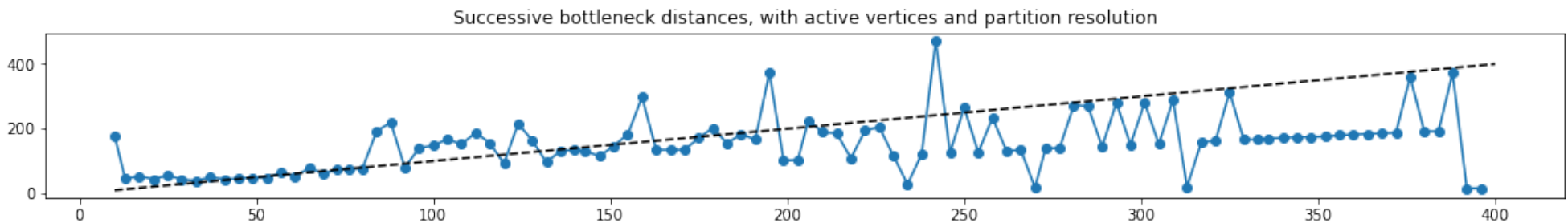
IV - Exemplos

Procuramos selecionar automaticamente a resolução.

1 - Calcular códigos de barras para muitas resoluções r_1, \dots, r_n .



2 - Para todos i , calcular a **distância bottleneck** entre os códigos de barras r_i e r_{i+1} . É referida como $d(r_i, r_{i+1})$.



Indo da resolução r_i à r_{i+1} , se distinguem dois valores de distância:

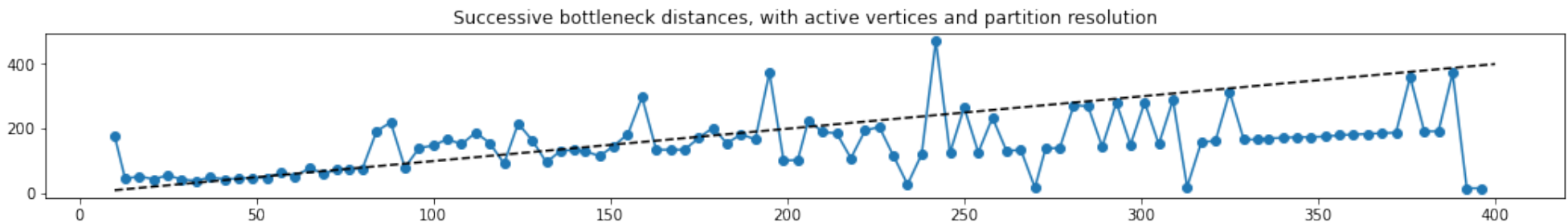
- se $d(r_i, r_{i+1}) \leq r_{i+1}$, então é uma **modificação local** do código de barras. Não é relevante.
- se $d(r_i, r_{i+1}) > r_{i+1}$, então é uma **mudança estrutural**. Isto significa que entramos em outro comportamento do grafo dinâmico.

Procuramos selecionar automaticamente a resolução.

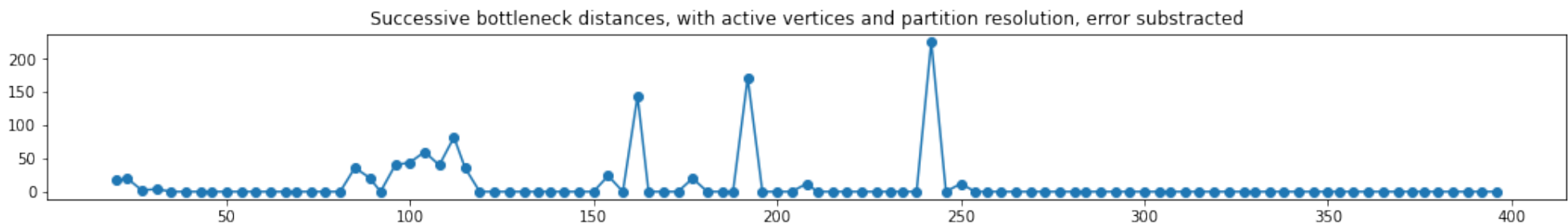
1 - Calcular códigos de barras para muitas resoluções r_1, \dots, r_n .



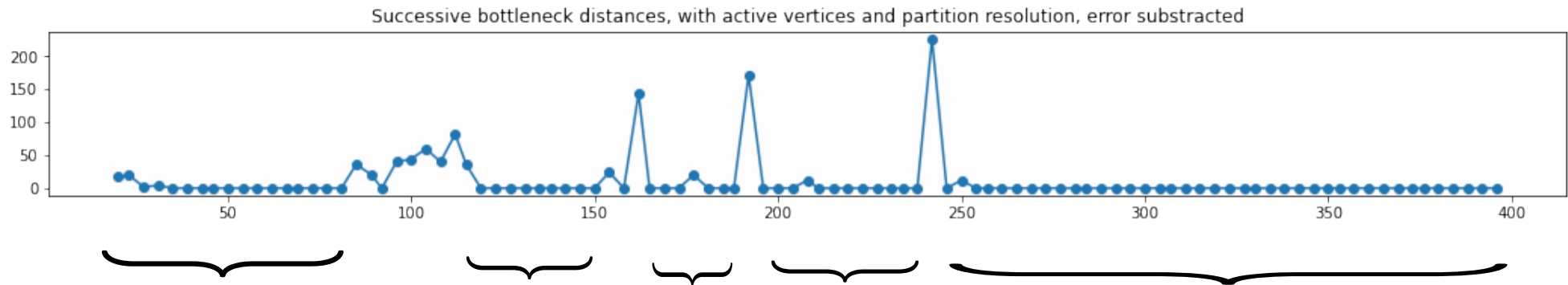
2 - Para todos i , calcular a **distância bottleneck** entre os códigos de barras r_i e r_{i+1} . É referida como $d(r_i, r_{i+1})$.



3 - Analisar a curva $i \mapsto \max(d(r_i, r_{i+1}) - r_{i+1}, 0)$.



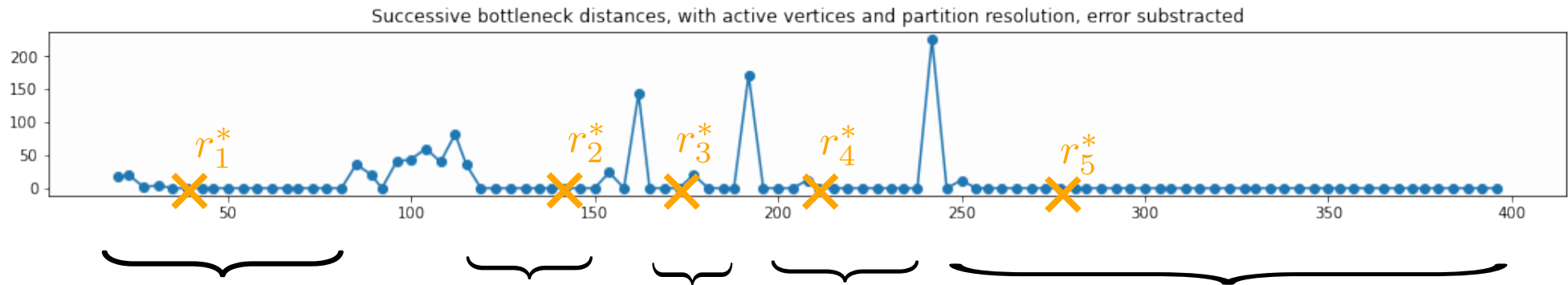
- 1 - Calcular códigos de barras para muitas resoluções r_1, \dots, r_n .
- 2 - Para todos i , calcular a **distância bottleneck** entre os códigos de barras r_i e r_{i+1} . É referida como $d(r_i, r_{i+1})$.
- 3 - Analisar a curva $i \mapsto \max(d(r_i, r_{i+1}) - r_{i+1}, 0)$.



Nessas faixas de valores, a distância entre os códigos de barras é zero. Assim, os gráficos correspondentes têm o mesmo comportamento.

Portanto, uma amostra representativa de cada comportamento é obtida selecionando uma resolução em cada uma das faixas de valores.

- 1 - Calcular códigos de barras para muitas resoluções r_1, \dots, r_n .
- 2 - Para todos i , calcular a **distância bottleneck** entre os códigos de barras r_i e r_{i+1} . É referida como $d(r_i, r_{i+1})$.
- 3 - Analisar a curva $i \mapsto \max(d(r_i, r_{i+1}) - r_{i+1}, 0)$.



Nessas faixas de valores, a distância entre os códigos de barras é zero. Assim, os gráficos correspondentes têm o mesmo comportamento.

Portanto, uma amostra representativa de cada comportamento é obtida selecionando uma resolução em cada uma das faixas de valores.

O usuário é dado os valores sugeridos $\{r_1^*, r_2^*, r_3^*, r_4^*, r_5^*\}$.

I - Homologia e homologia persistente

II - Código de barras decorado

III - Sugestão de resolução

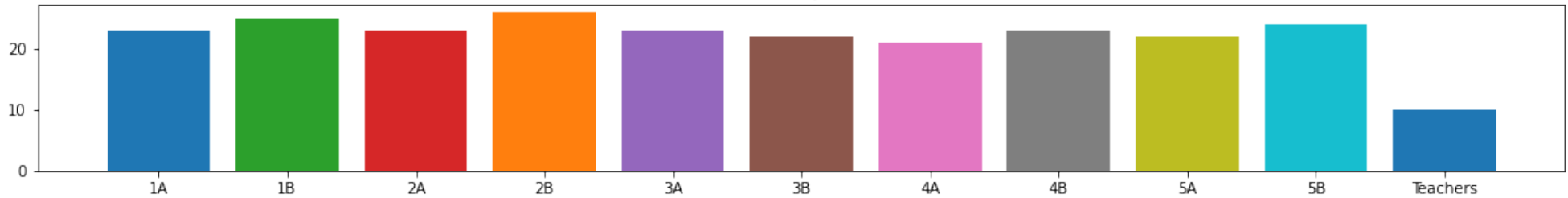
IV - Exemplos

Primary school dataset

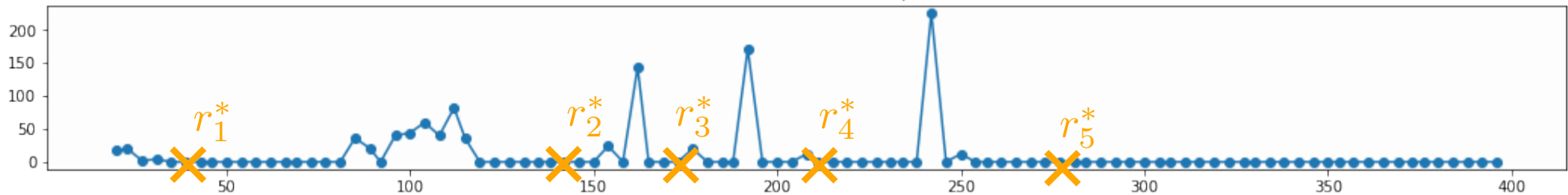
17/21 (1/6)

Este conjunto de dados representa as conexões entre os alunos de uma escola primária durante uma semana.

Clusters repartition



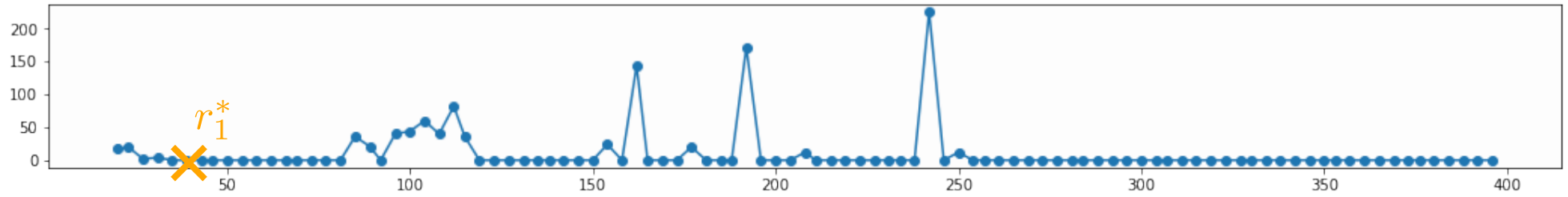
Successive bottleneck distances, with active vertices and partition resolution, error subtracted



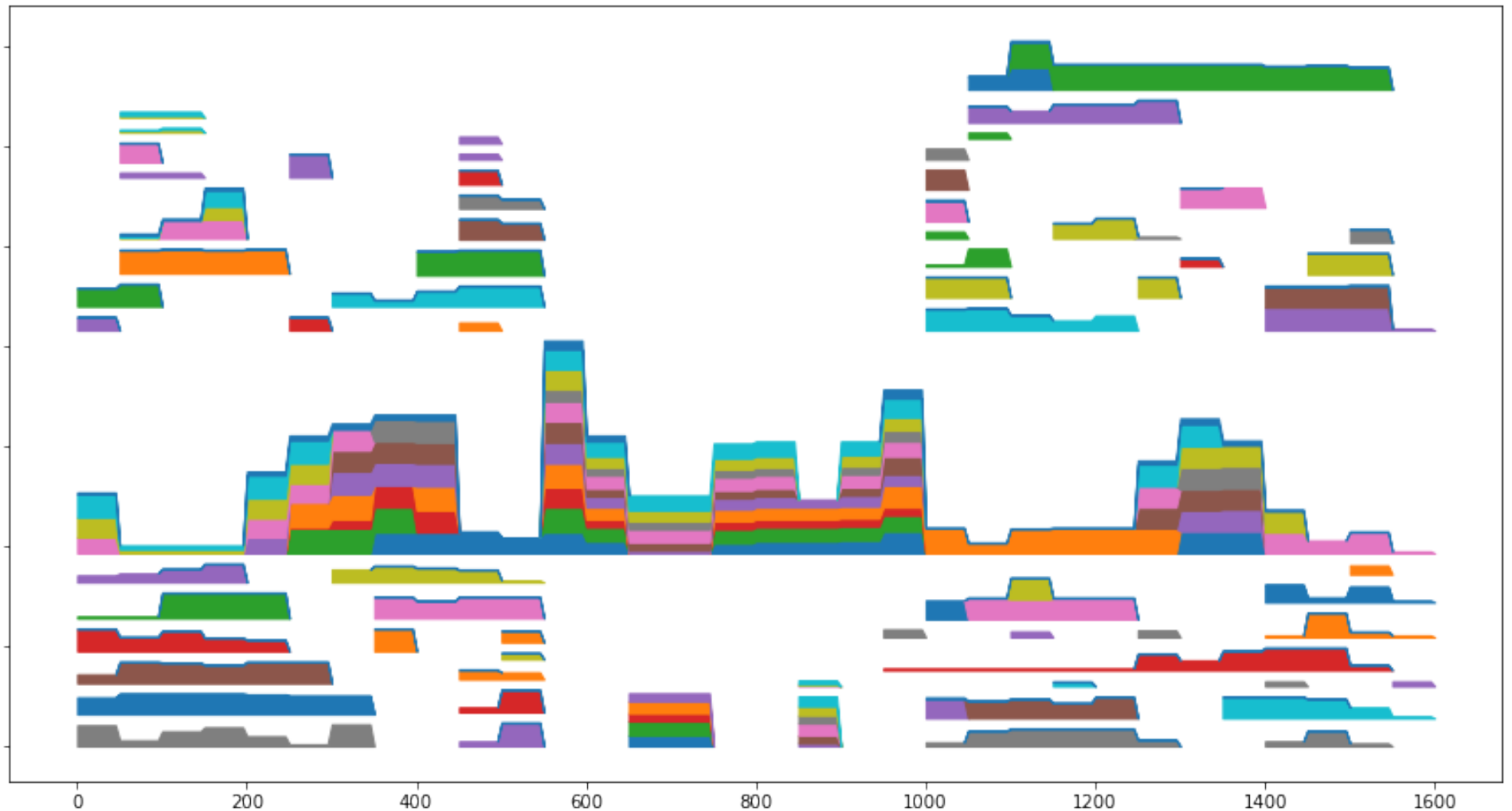
Primary school dataset

17/21 (2/6)

Successive bottleneck distances, with active vertices and partition resolution, error subtracted



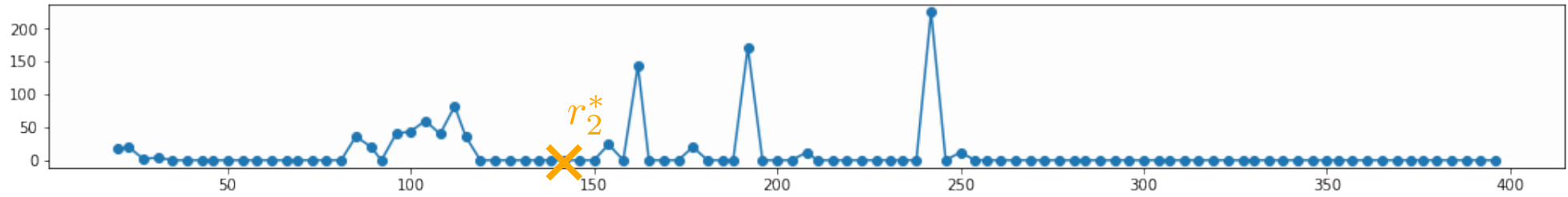
Resolution = 50



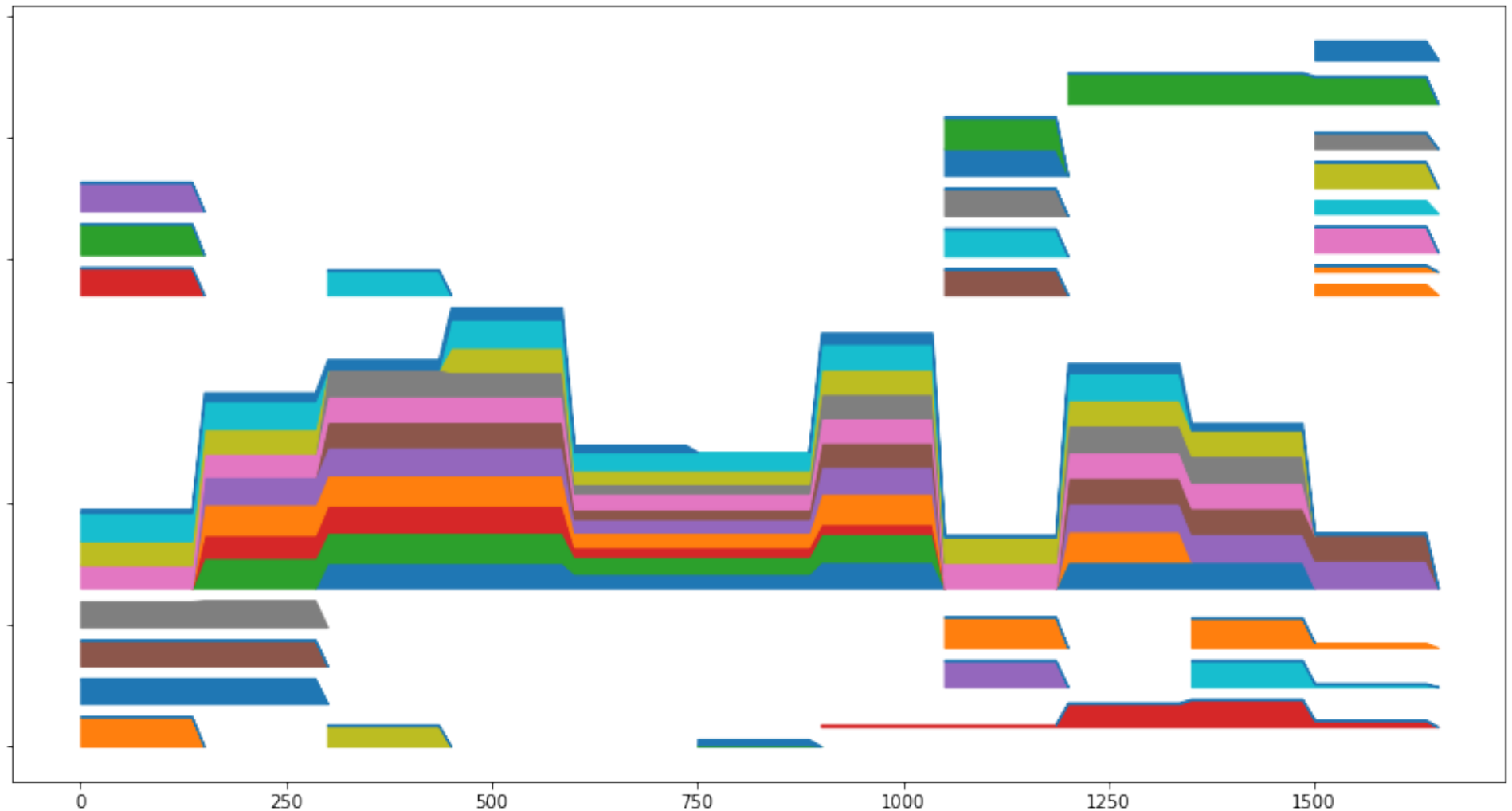
Primary school dataset

17/21 (3/6)

Successive bottleneck distances, with active vertices and partition resolution, error subtracted



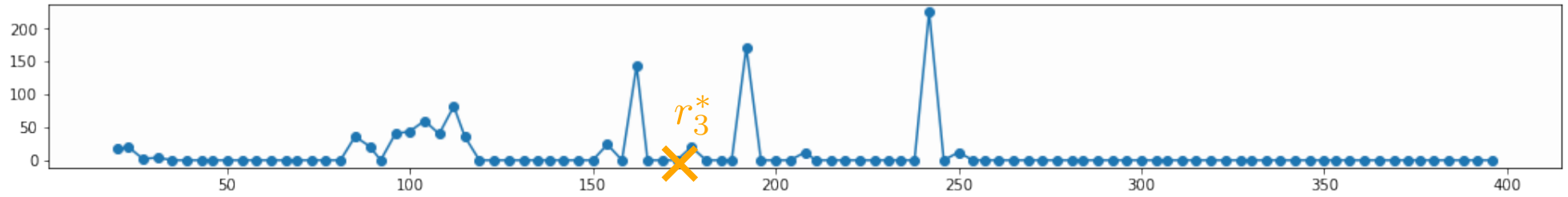
Resolution = 150



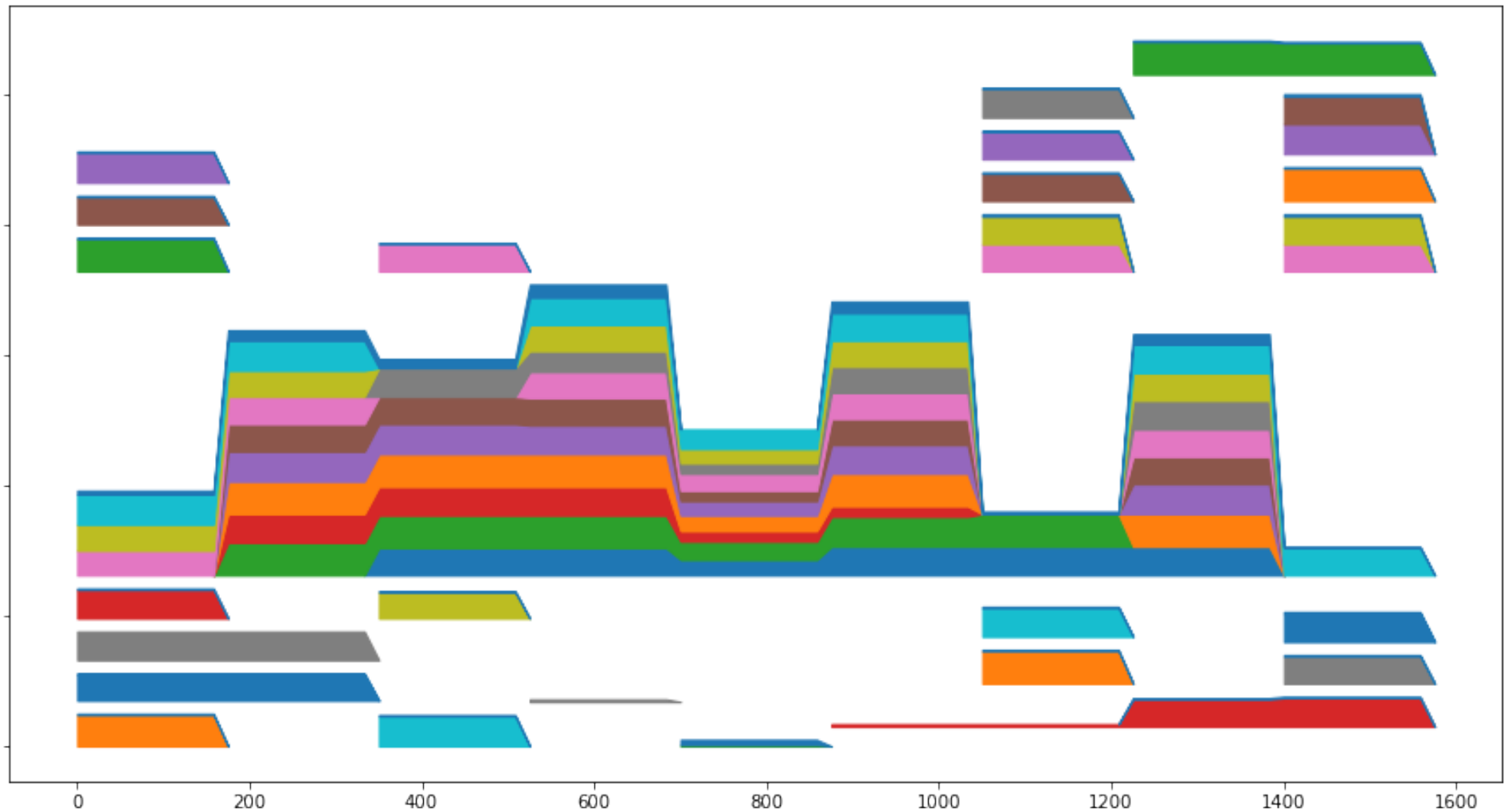
Primary school dataset

17/21 (4/6)

Successive bottleneck distances, with active vertices and partition resolution, error subtracted

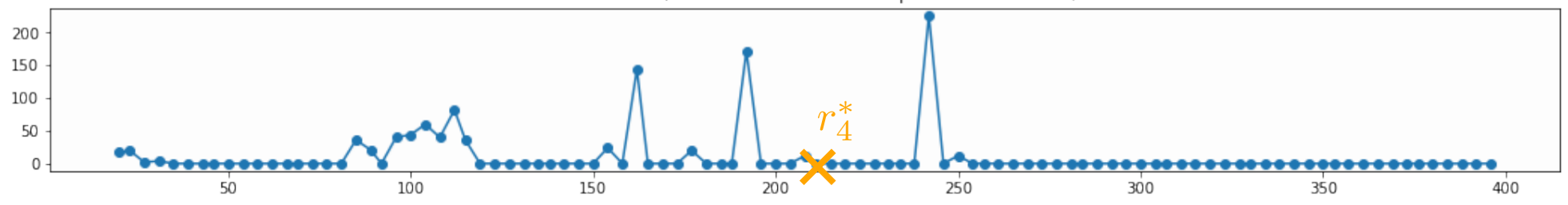


Resolution = 175

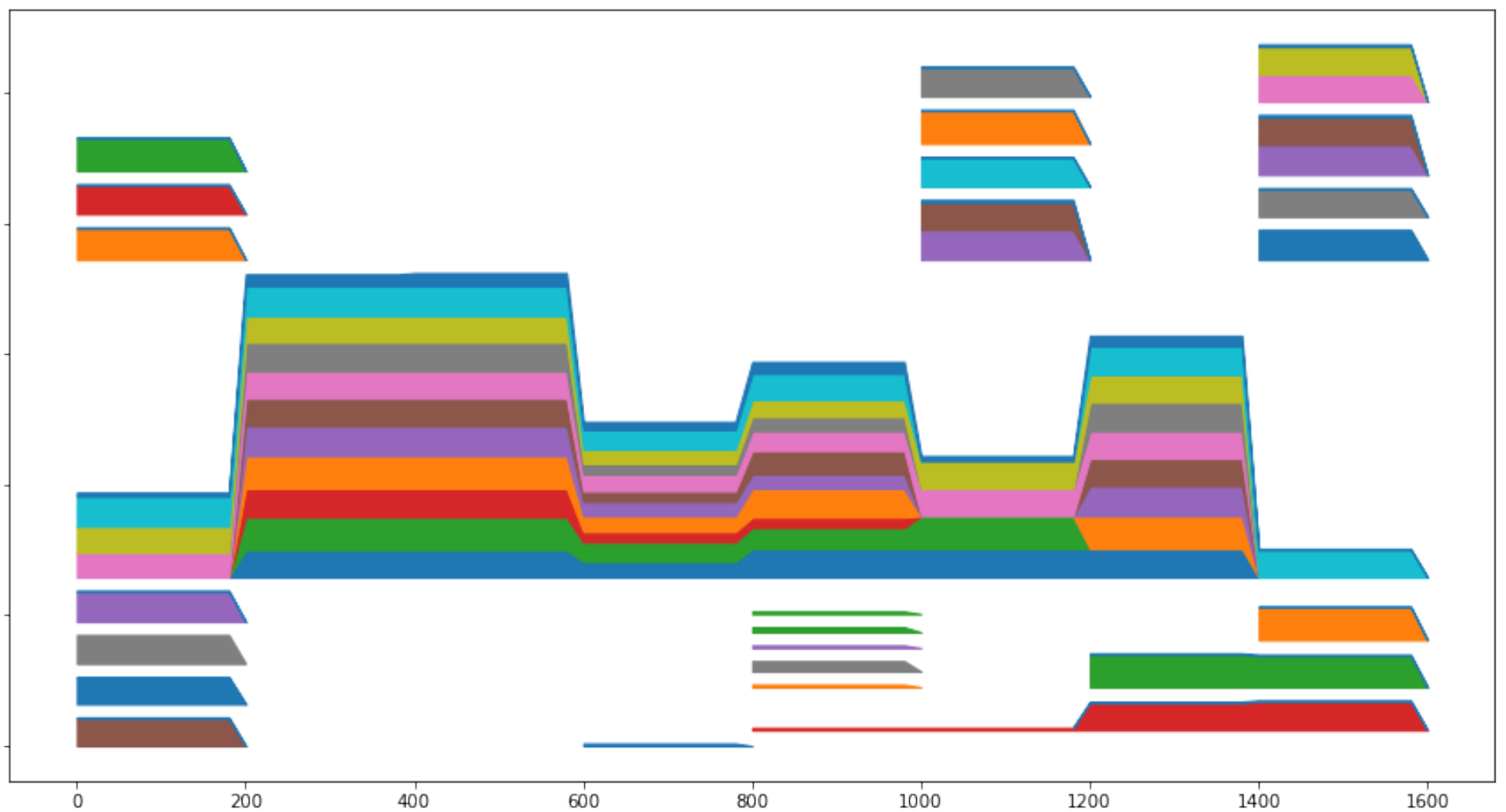


Primary school dataset

Successive bottleneck distances, with active vertices and partition resolution, error subtracted

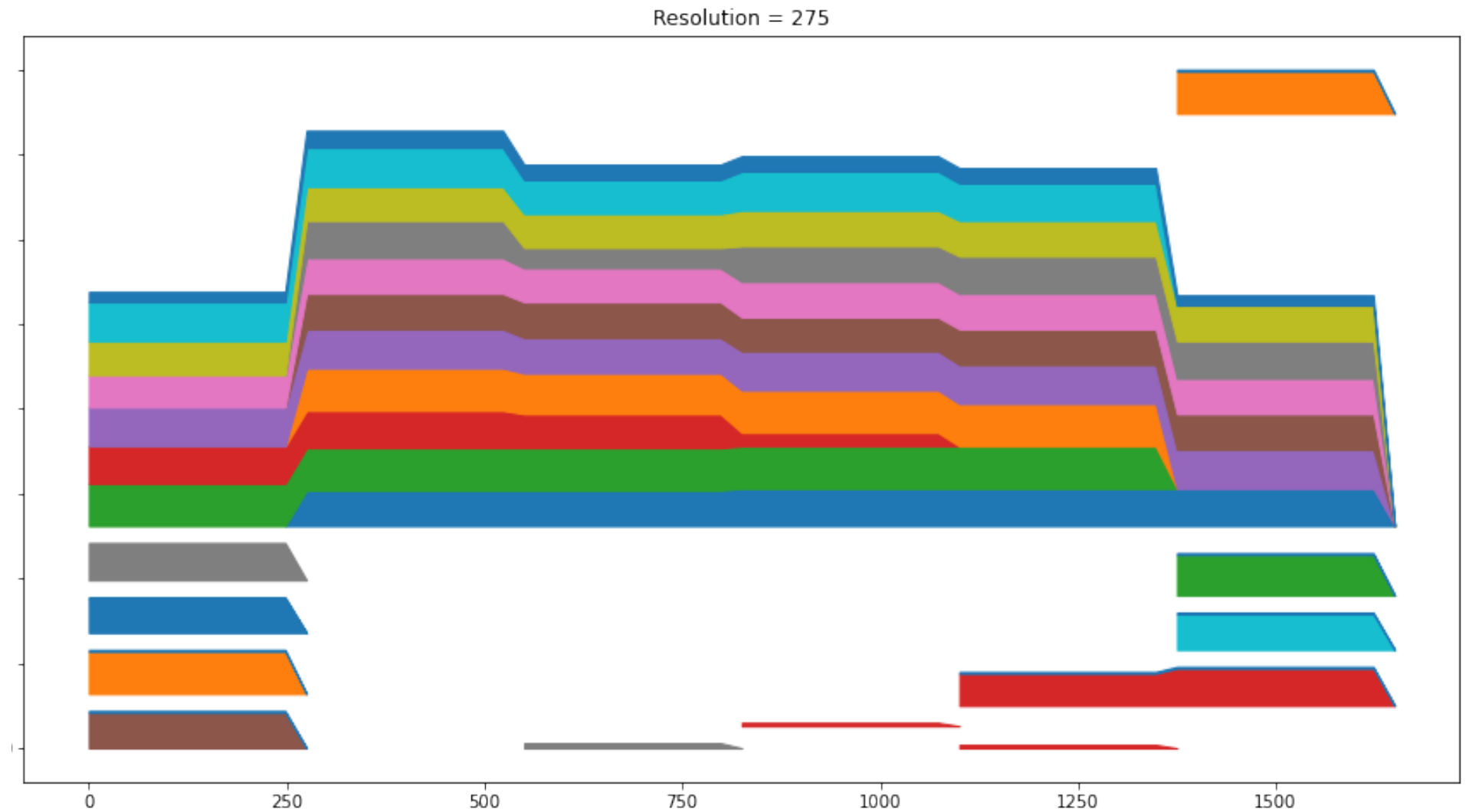
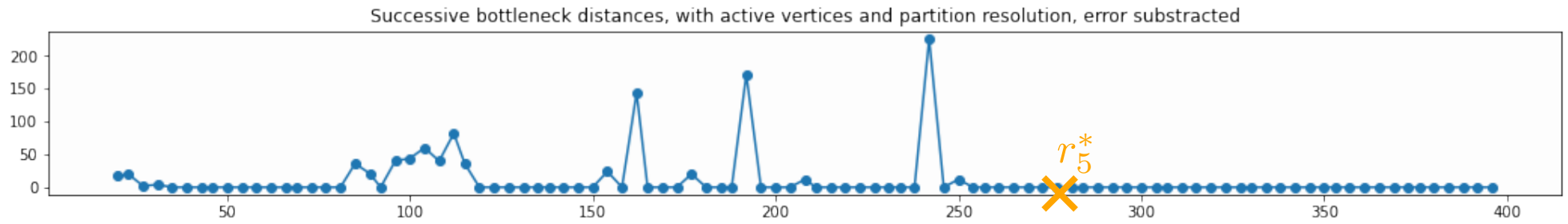


Resolution = 200



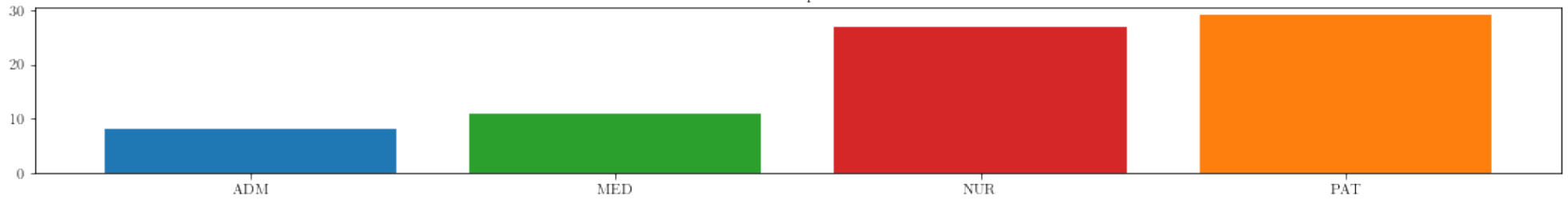
Primary school dataset

17/21 (6/6)

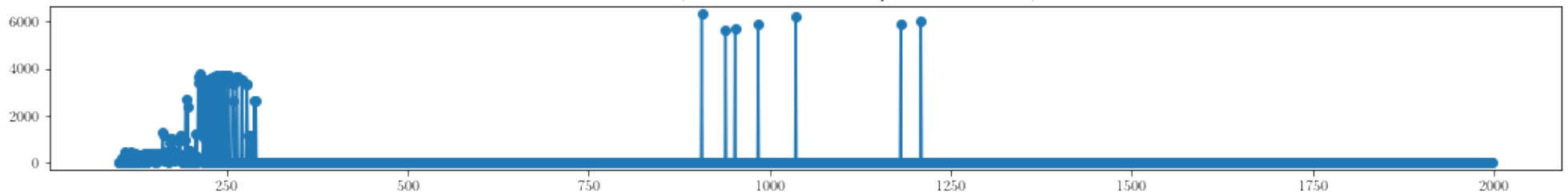


Conexões entre os membros de um hospital.

Clusters repartition

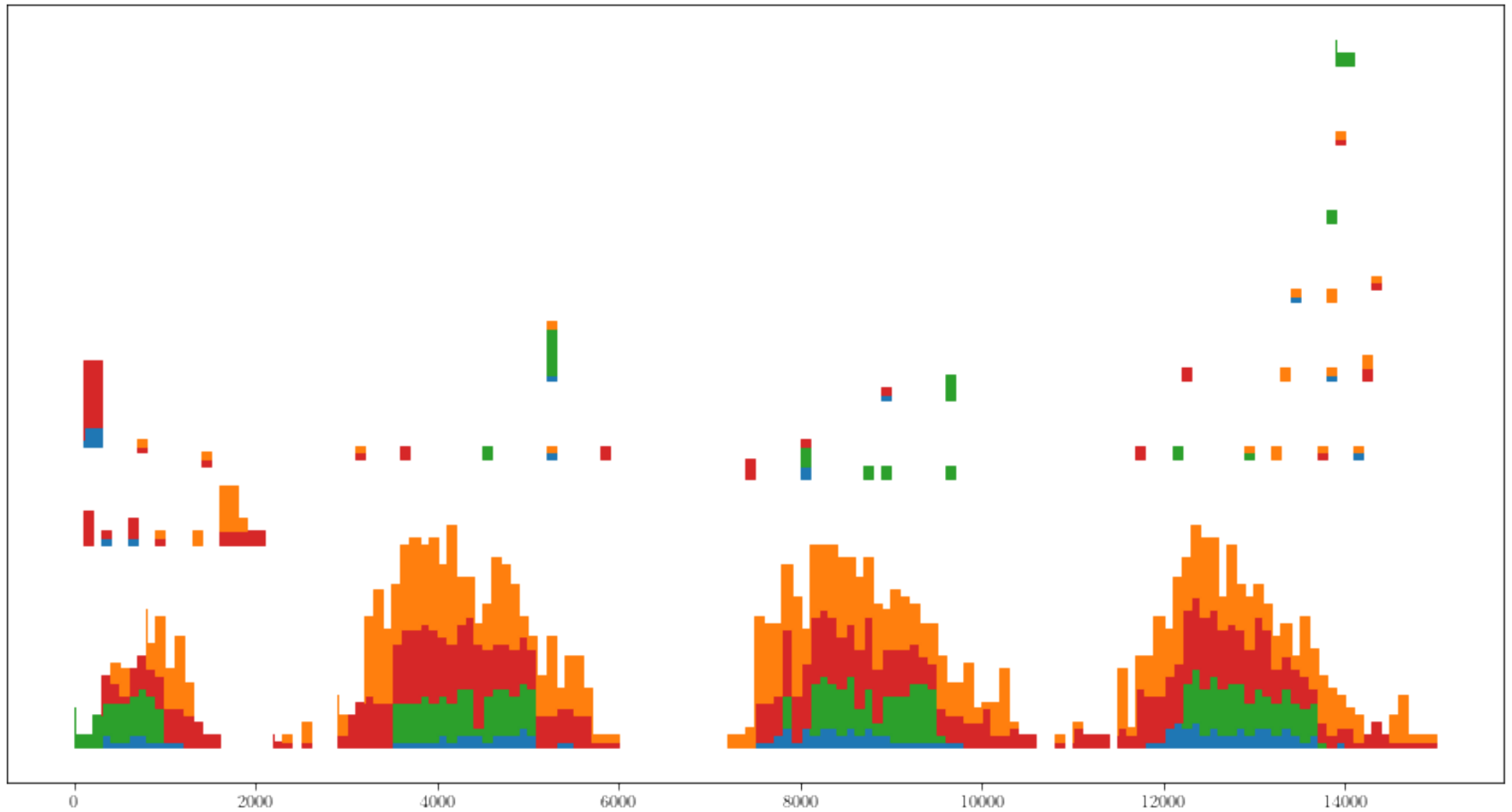
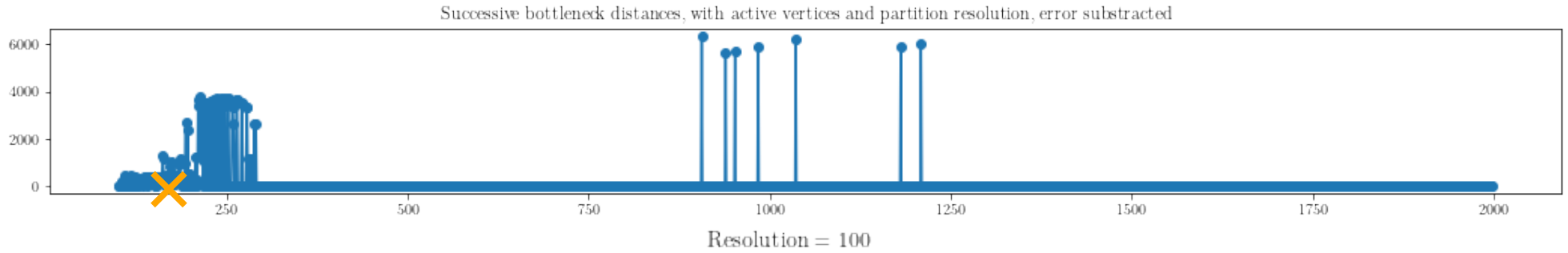


Successive bottleneck distances, with active vertices and partition resolution, error substracted



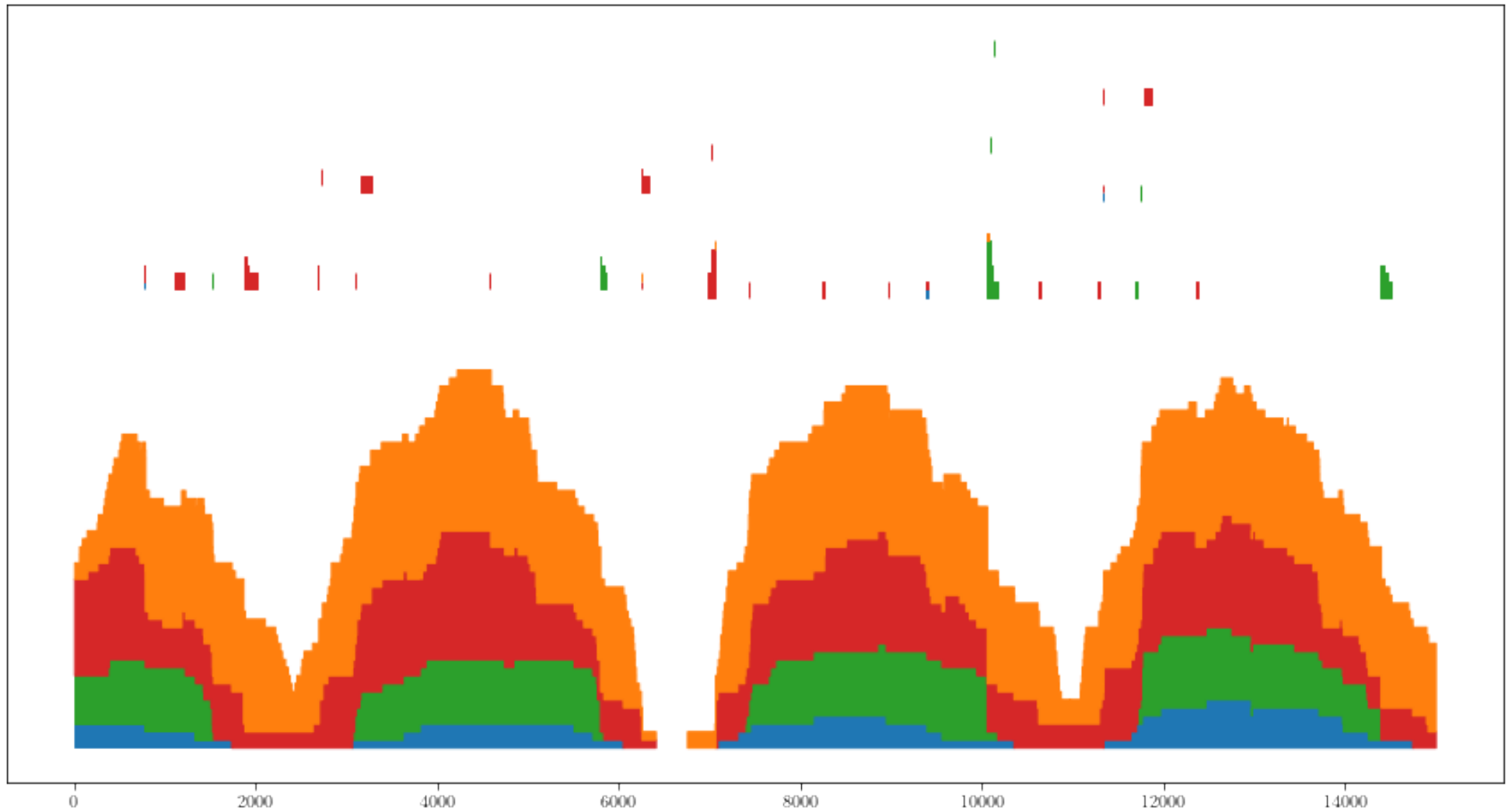
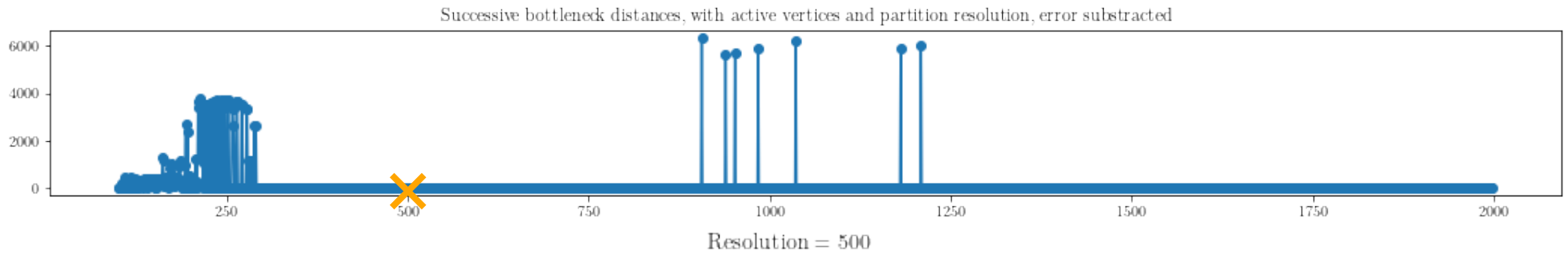
Hospital

18/21 (2/8)



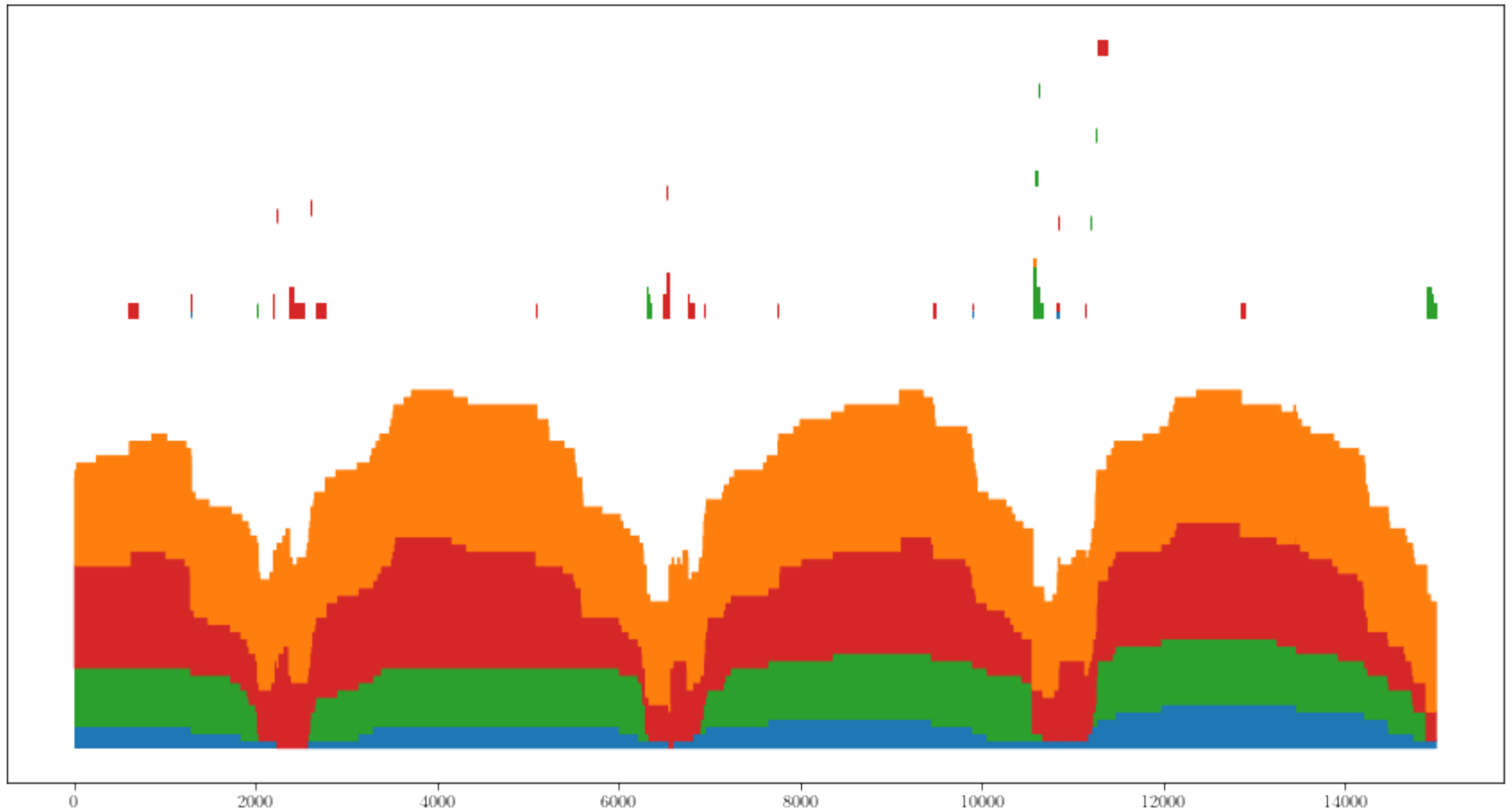
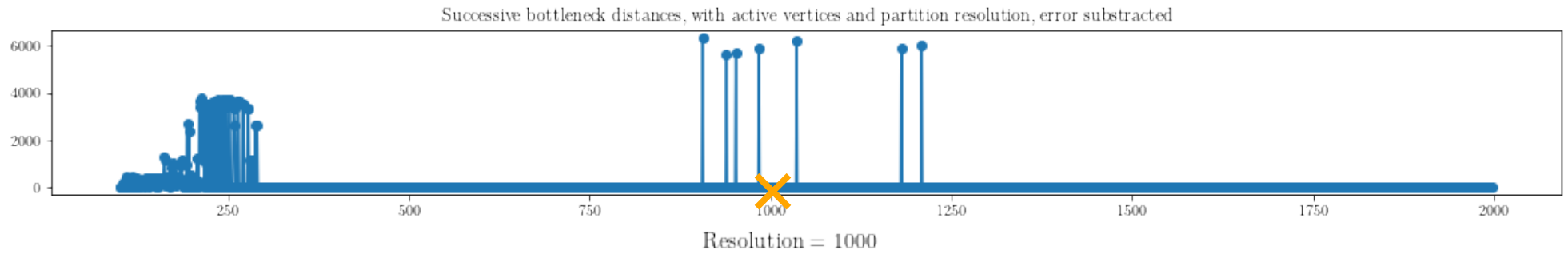
Hospital

18/21 (3/8)

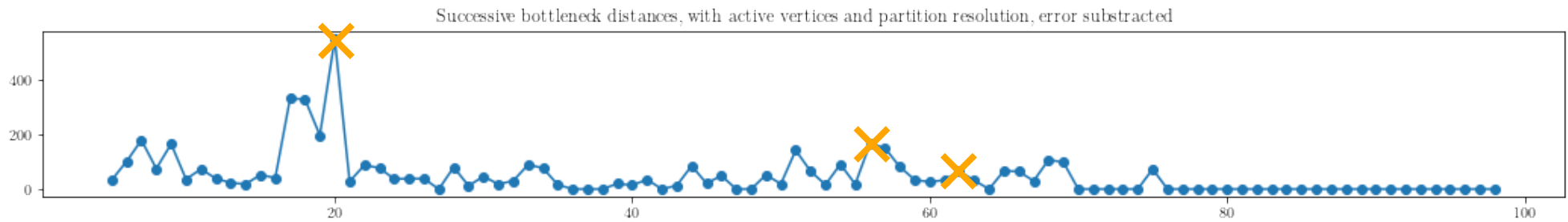


Hospital

18/21 (4/8)

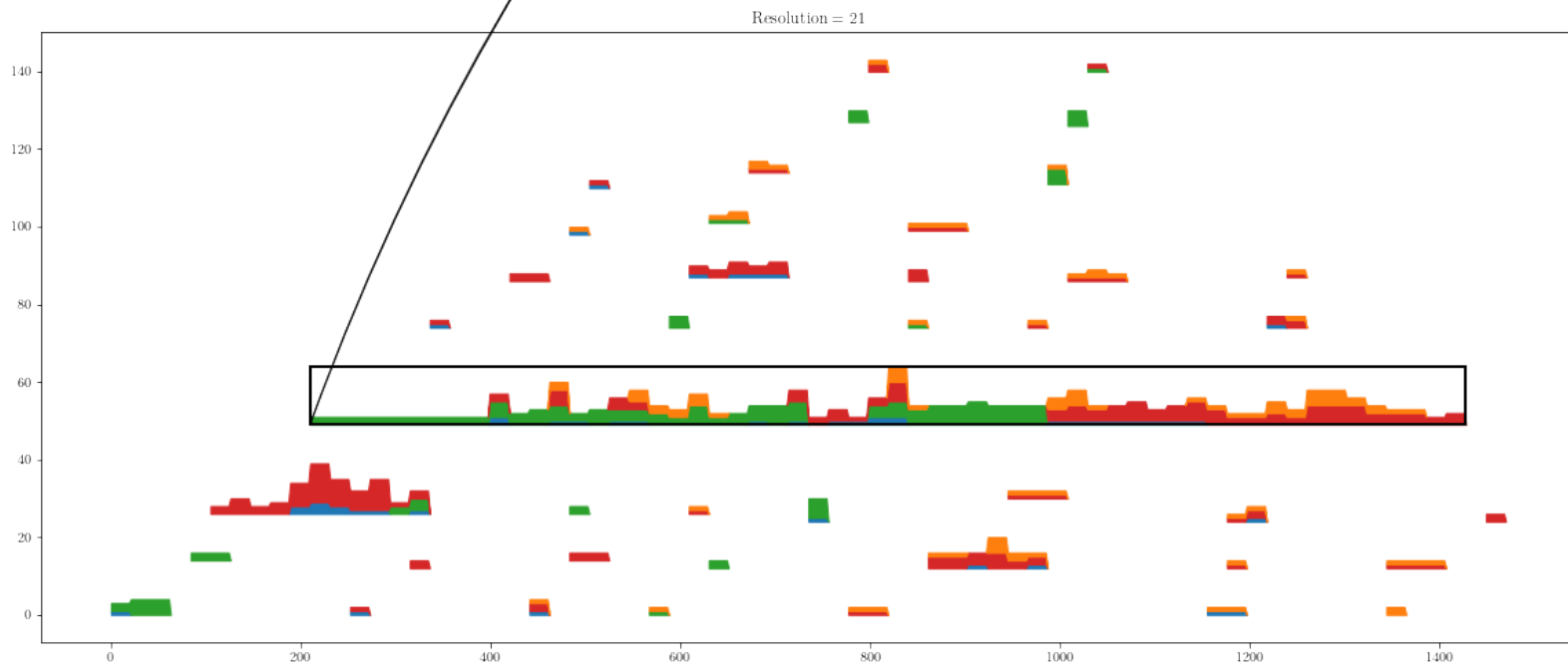
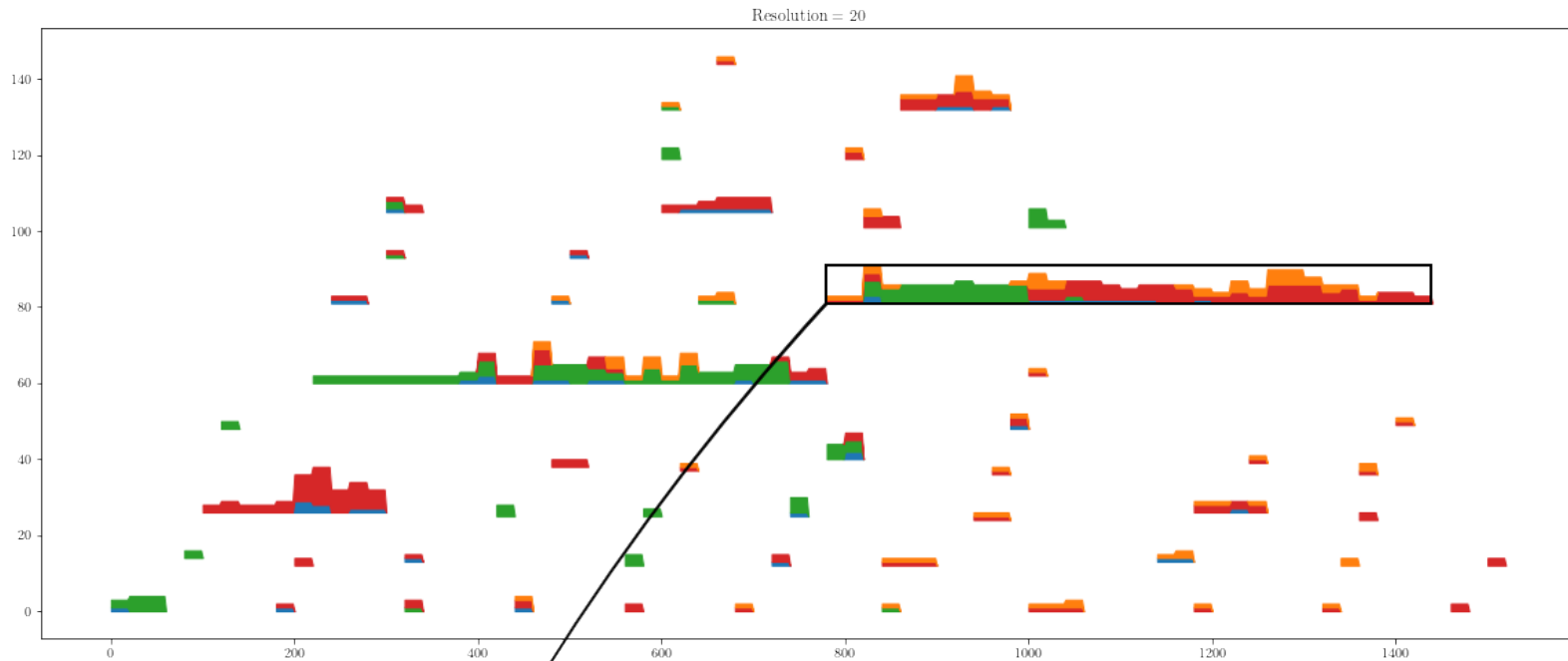


Podemos dar zoom no primeiro dia:



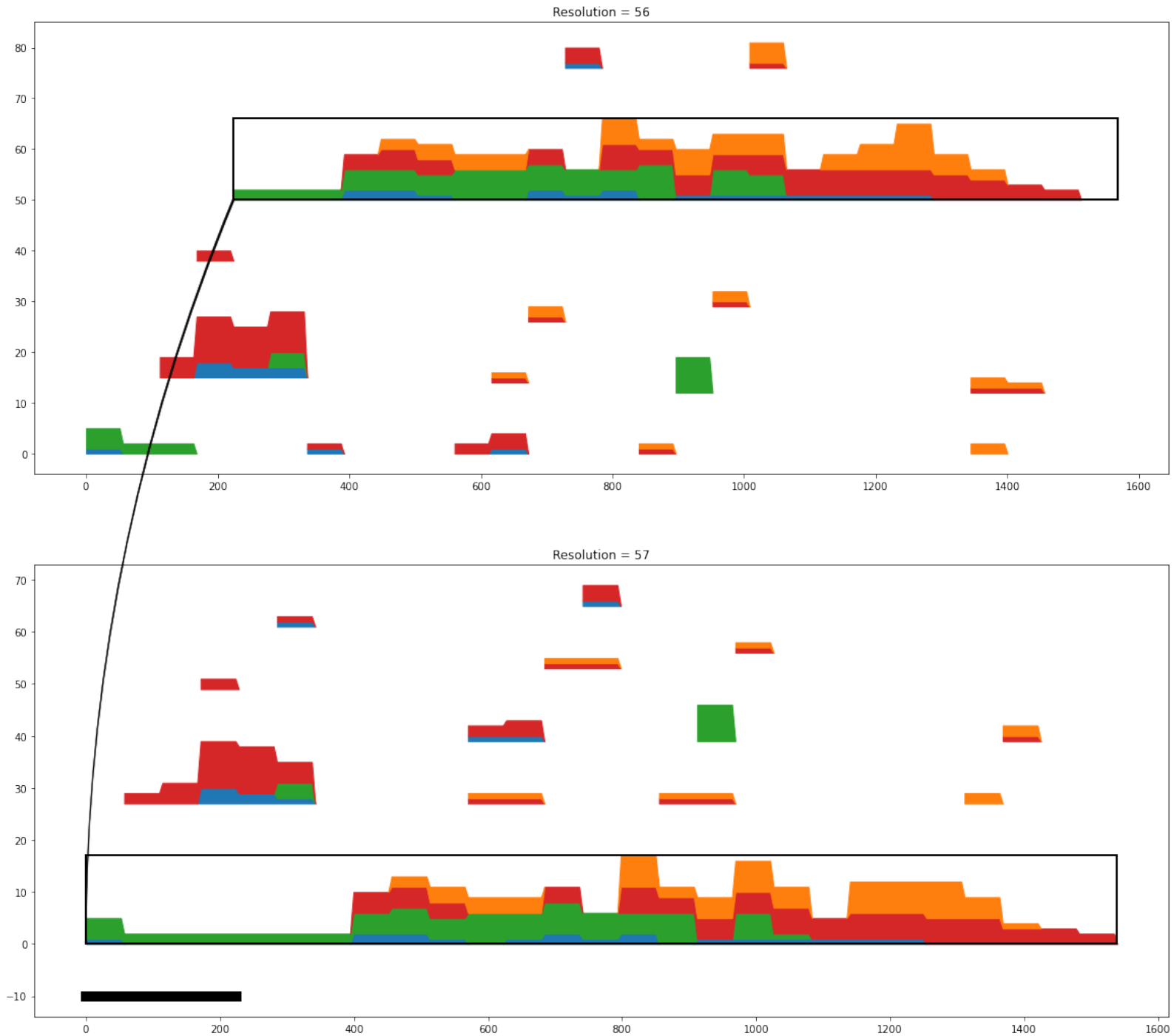
Hospital

18/21 (6/8)



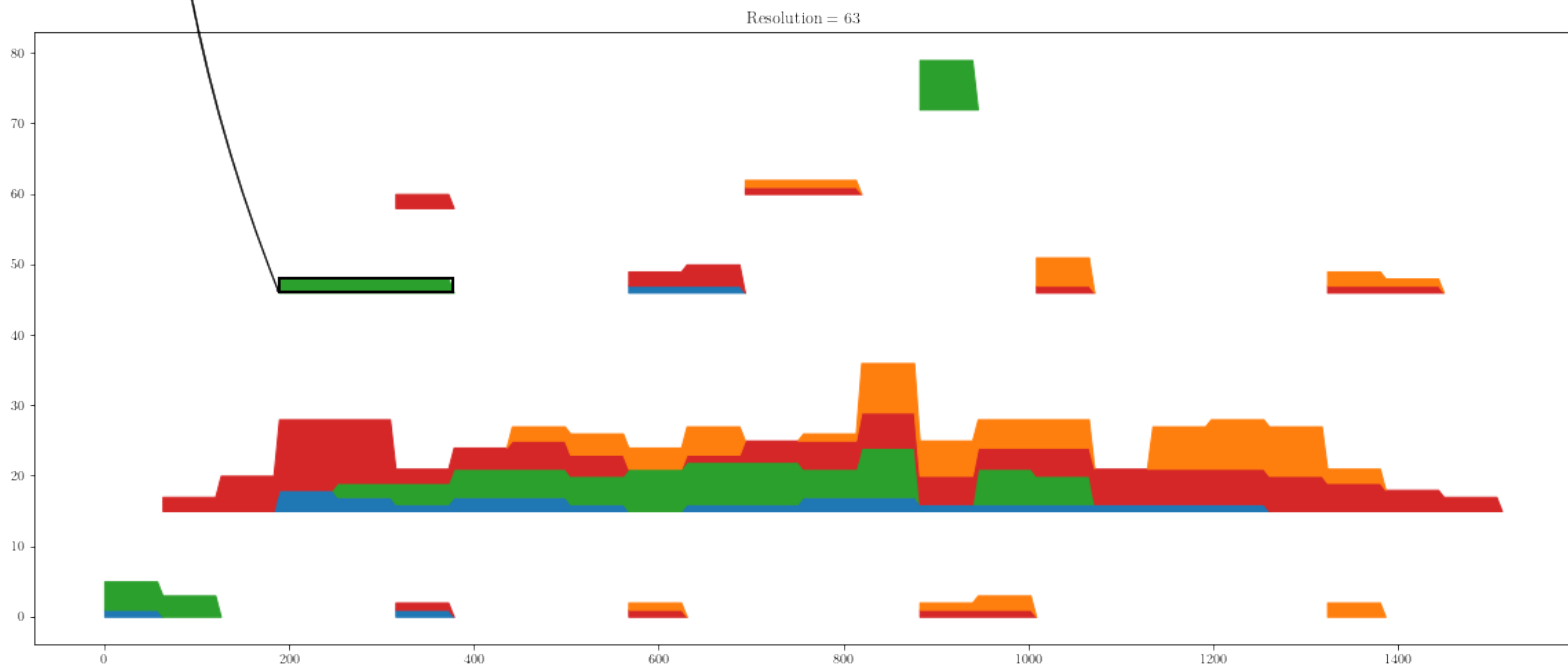
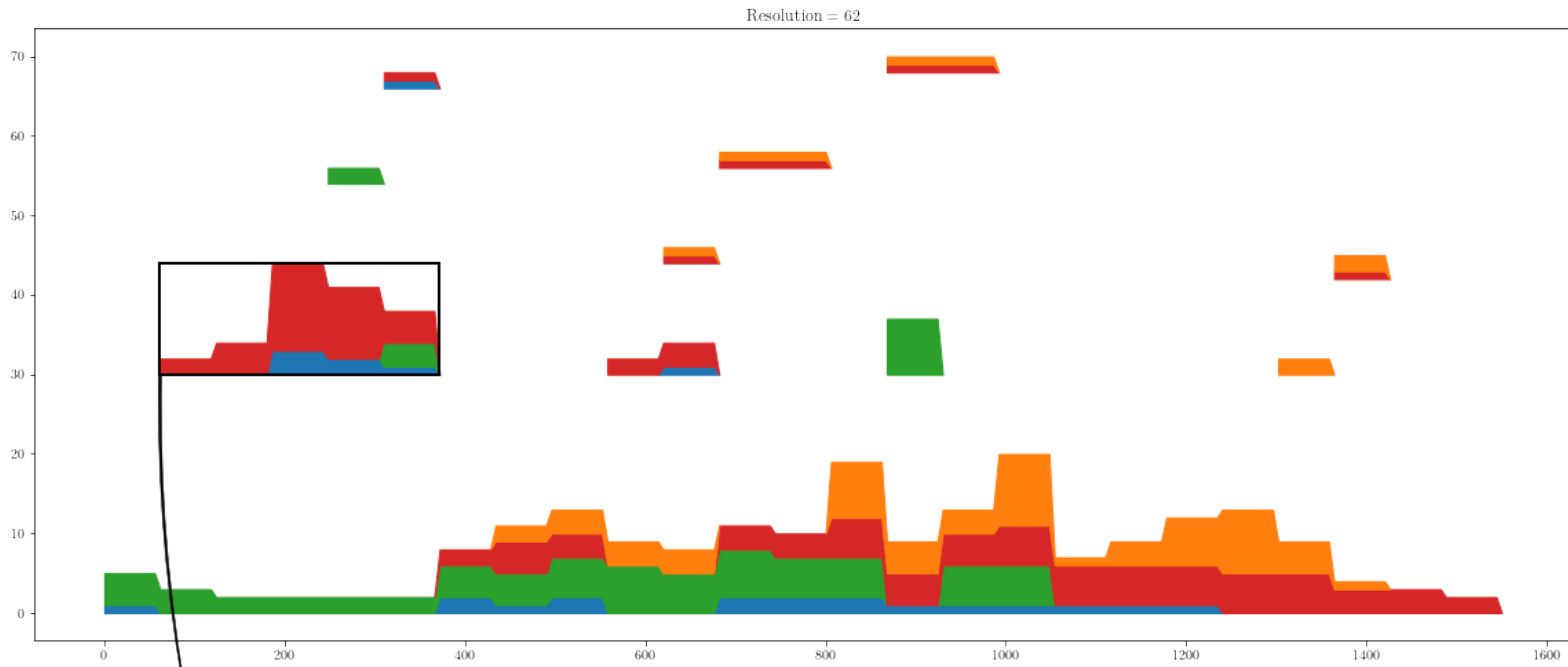
Hospital

18/21 (7/8)



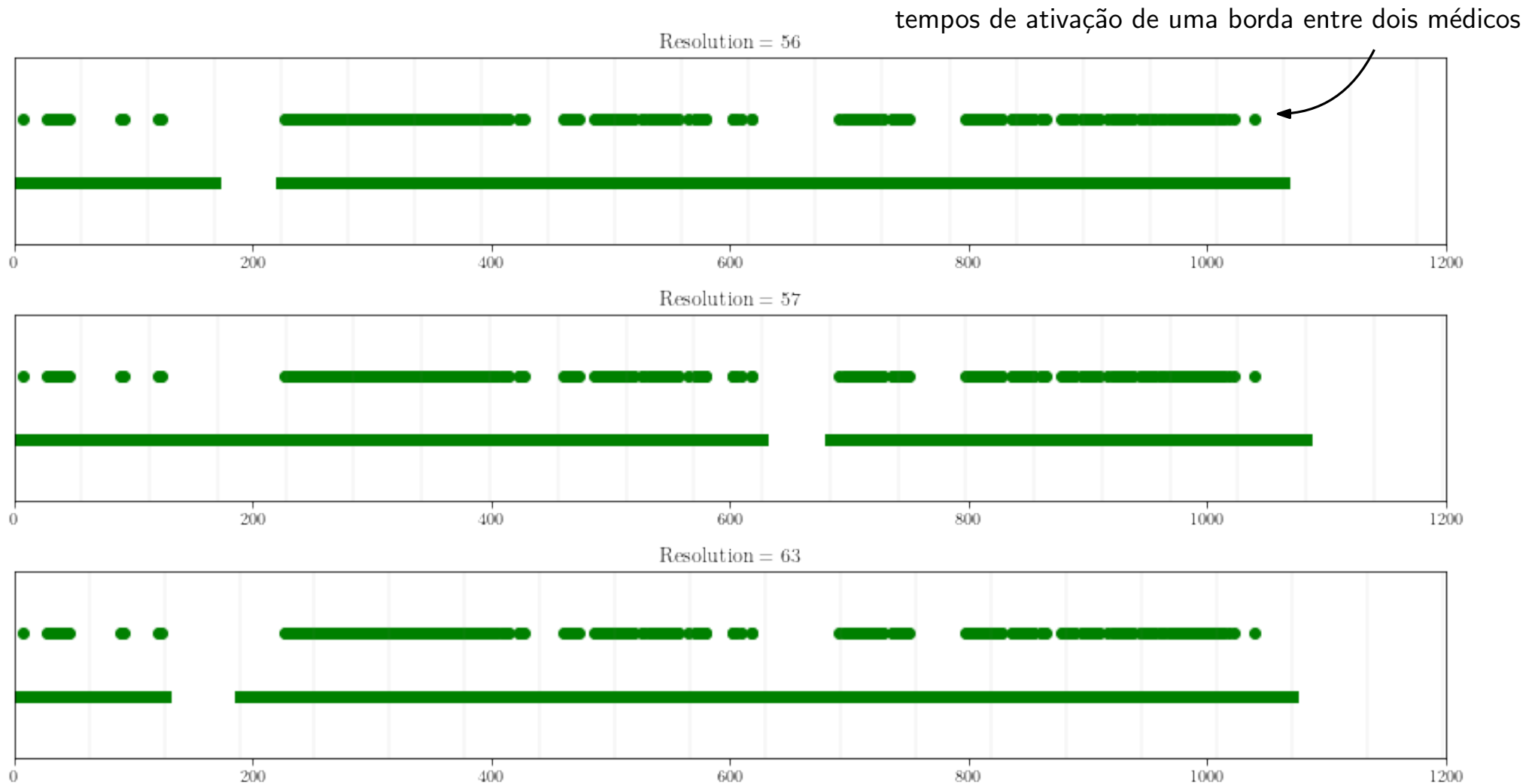
Hospital

18/21 (8/8)



Instabilidade do corte em sub-intervalos 19/21

Da resolução 56 a 57, depois 62 a 63, observa-se que uma grande barra verde se forma e depois se recorta.



Para evitar este fenómeno, outro método de corte pode ser usado.

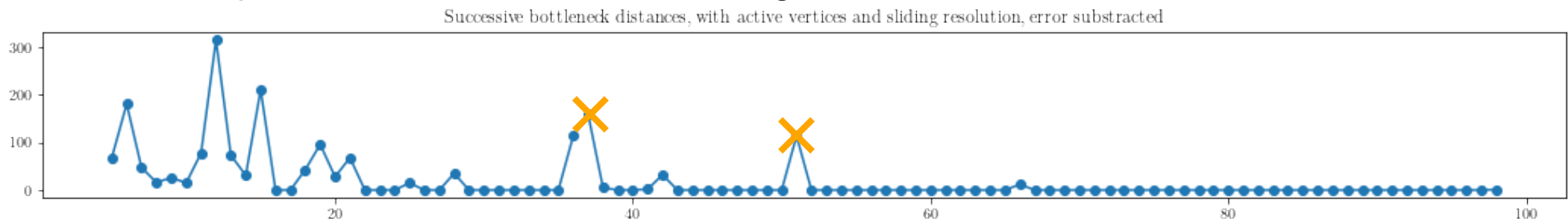
Corte em janela móvel (sliding window)_{20/21 (1/3)}

Um grafo temporal consiste nos dados de um grafo G e uma coleção de pares (v, t) , onde v é uma aresta de G e t um número real em $[0, T]$.

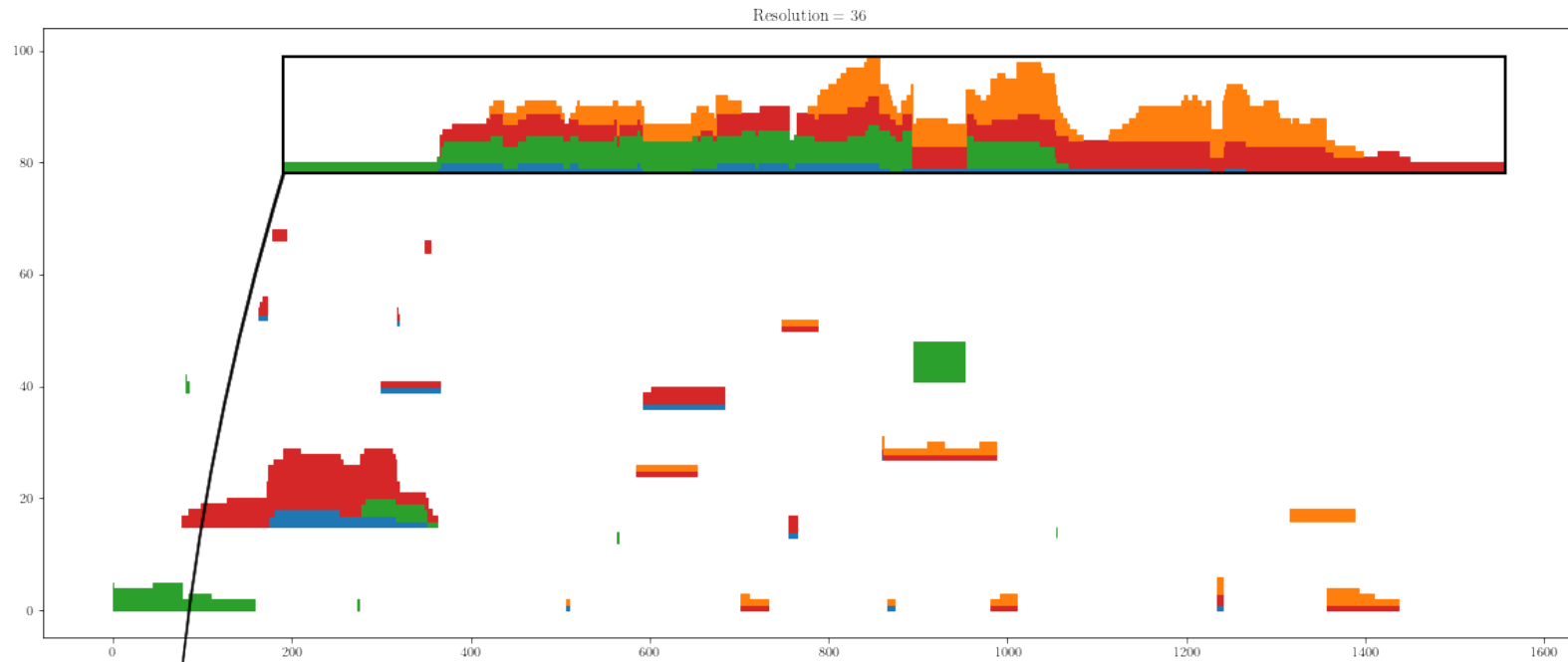
Corte em partição: Dada uma resolução r , cortamos o intervalo $[0, T]$ em $n = \lceil T/r \rceil$ subintervalos $[kr, (k+1)r]$, e construímos grafos $\{G_k | k = 0, \dots, n-1\}$.

Corte em janela móvel: Dada uma resolução r , cada aresta temporal (v, t) é associada a uma janela de ativação $[t-r, t+r]$. Cortamos o intervalo $[0, T]$ em T subintervalos $[k, (k+1)]$, e construímos grafos $\{G_k | k = 0, \dots, T-1\}$.

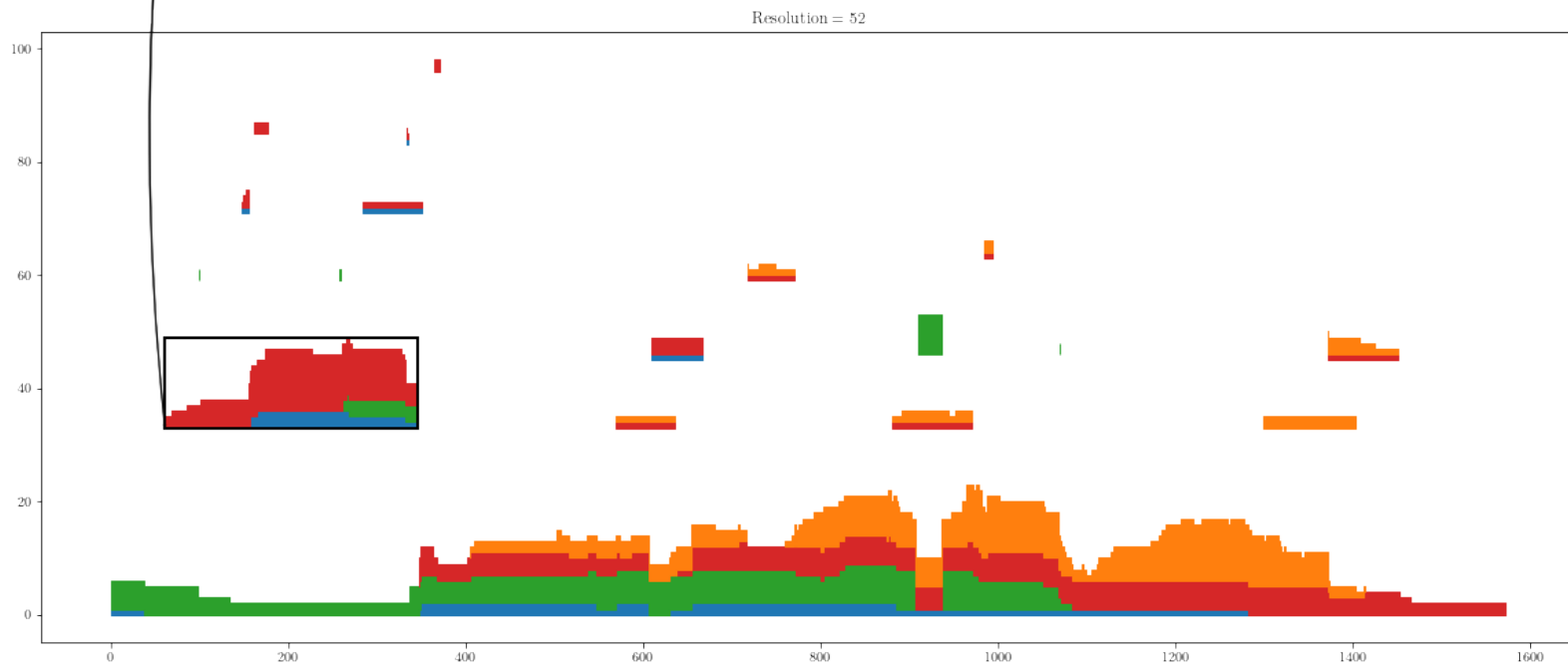
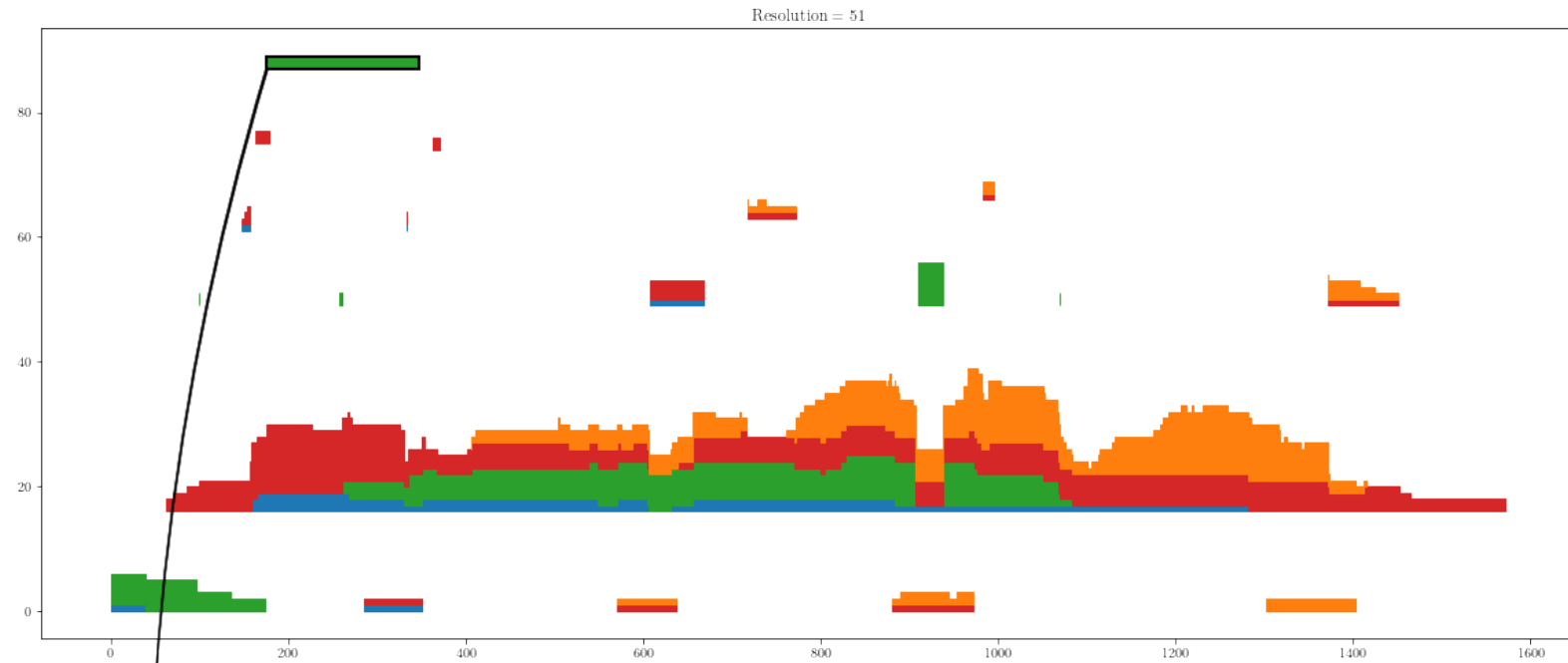
Como antes, podemos definir a curva de sugestão.



Corte em janela móvel (sliding window) 20/21 (2/3)

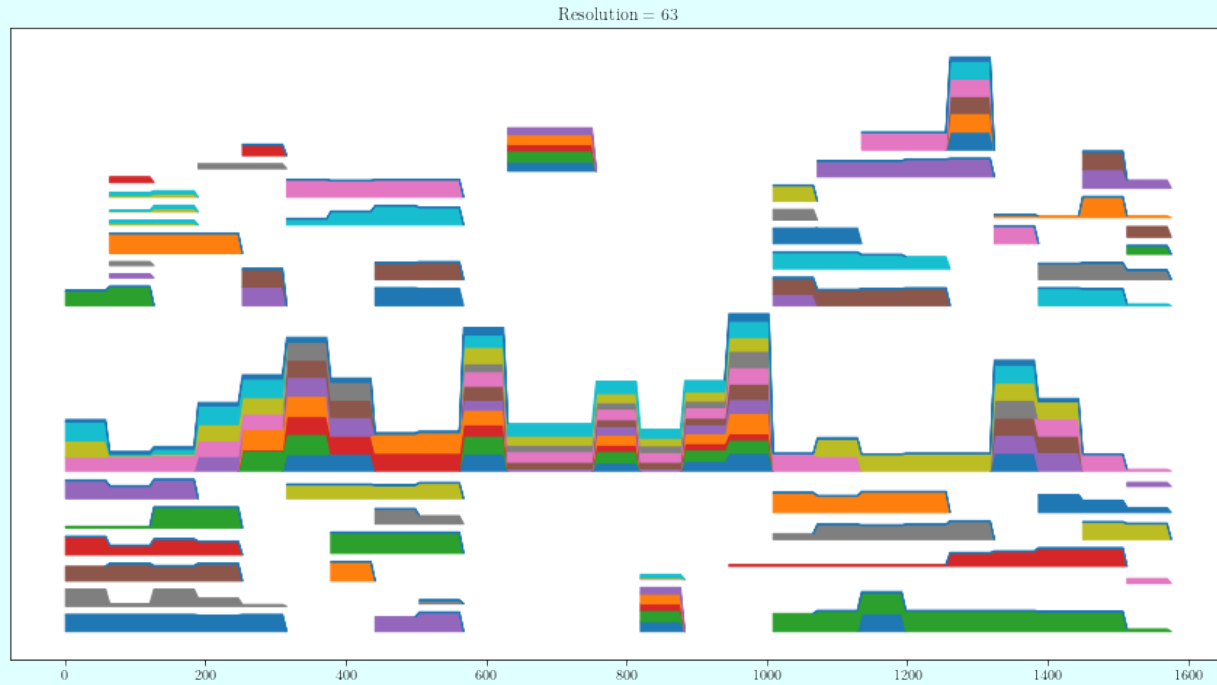


Corte em janela móvel (sliding window) 20/21 (3/3)

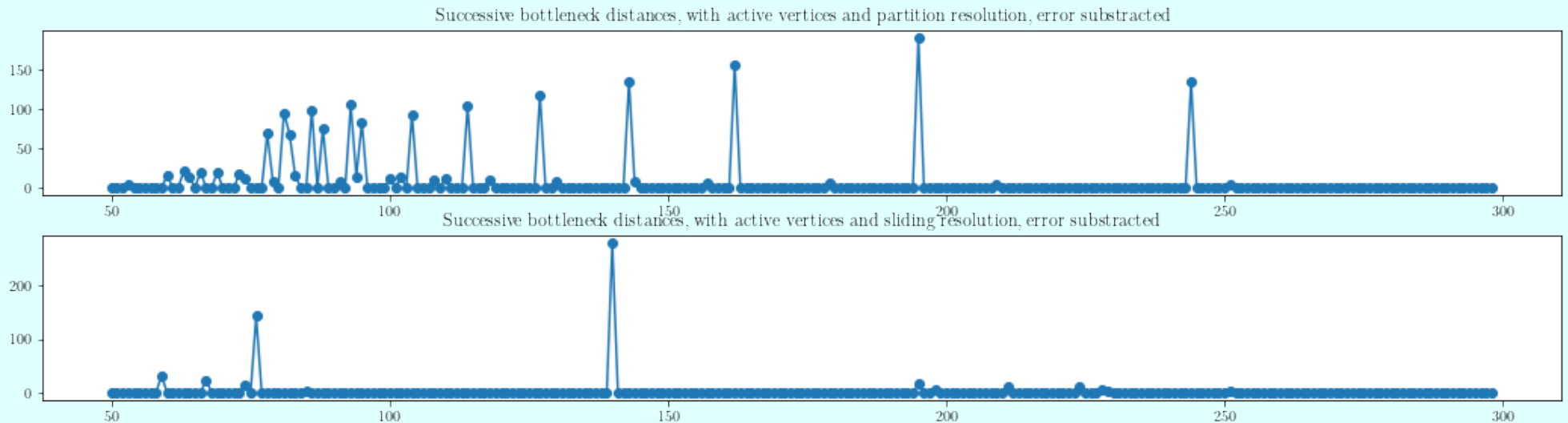


Conclusão

Propusemos uma nova visualização dos grafos temporais.

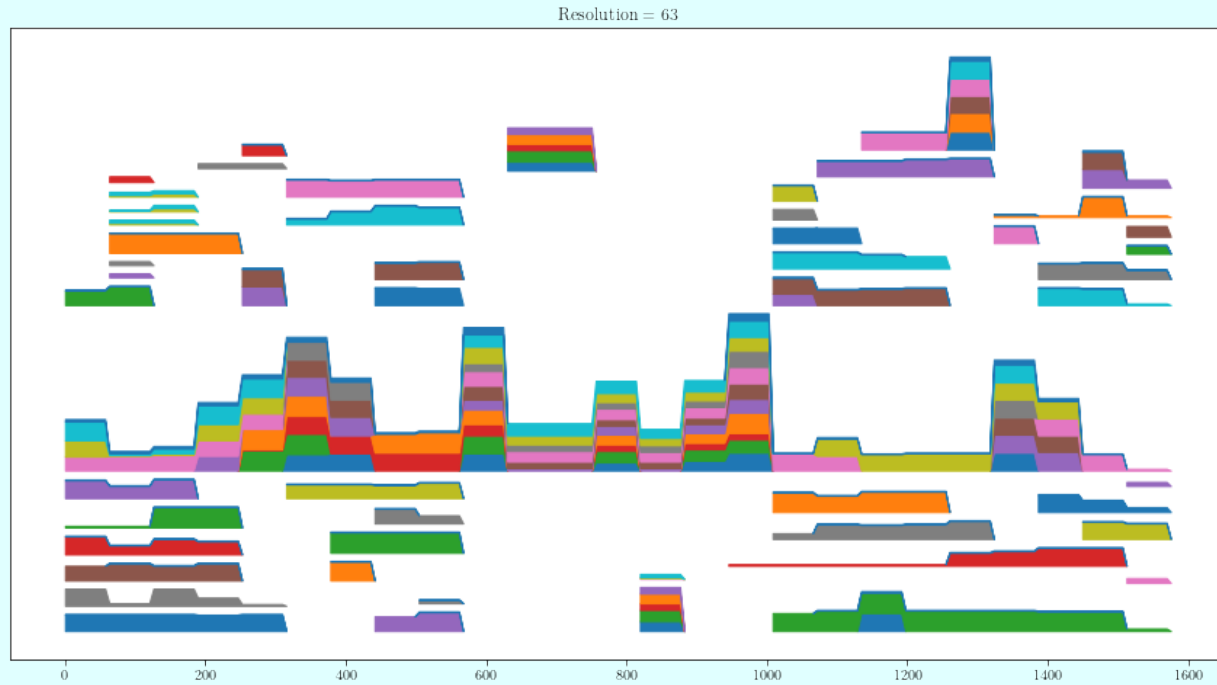


Sugestões de resolução através da curva de distâncias bottleneck sucessivas, com partição ou janela móvel.

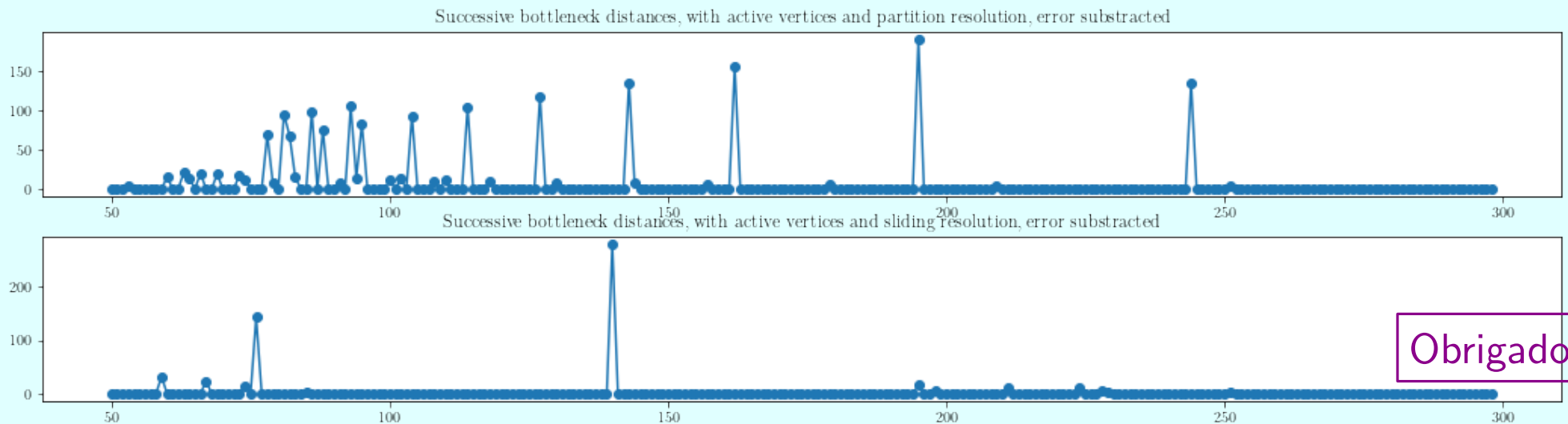


Conclusão

Propusemos uma nova visualização dos grafos temporais.



Sugestões de resolução através da curva de distâncias bottleneck sucessivas, com partição ou janela móvel.



Obrigado!