Graph-based social media story linking

TRECVID 2018 - Social-media video story-telling linking Task

Goncalo Marcelino, Joao Magalhaes

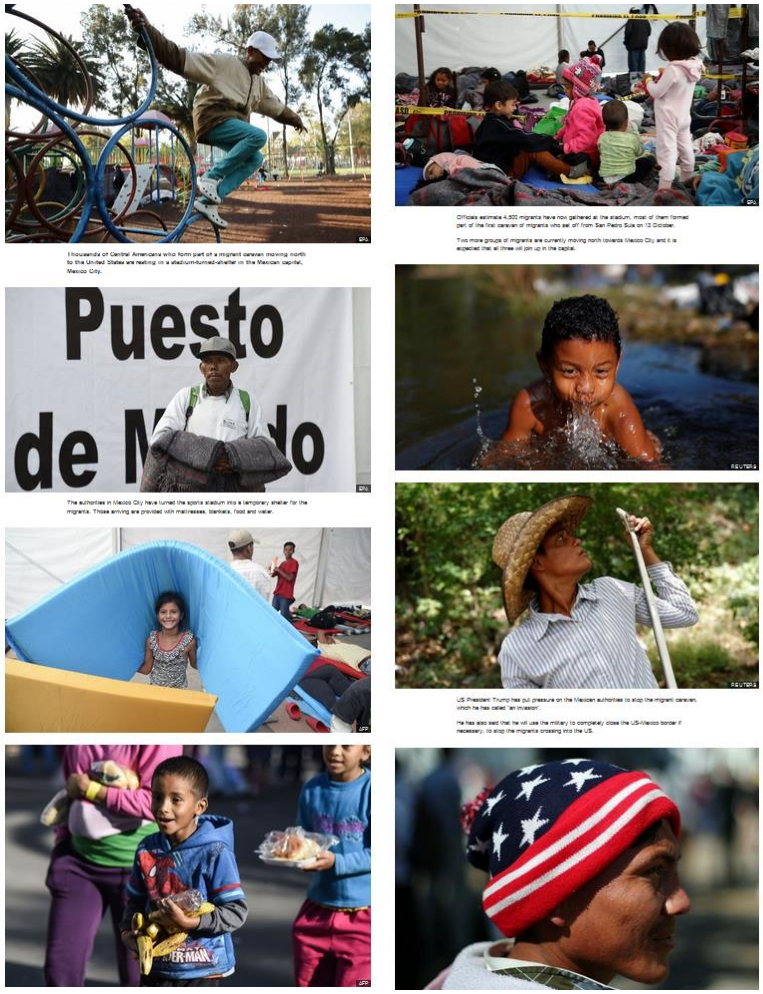
NOVA LINCS - Faculdade de Ciências e Tecnologia Universidade NOVA Lisboa, Caparica, Portugal

goncalo.bfm@gmail.com, jmag@fct.unl.pt

Context and motivation

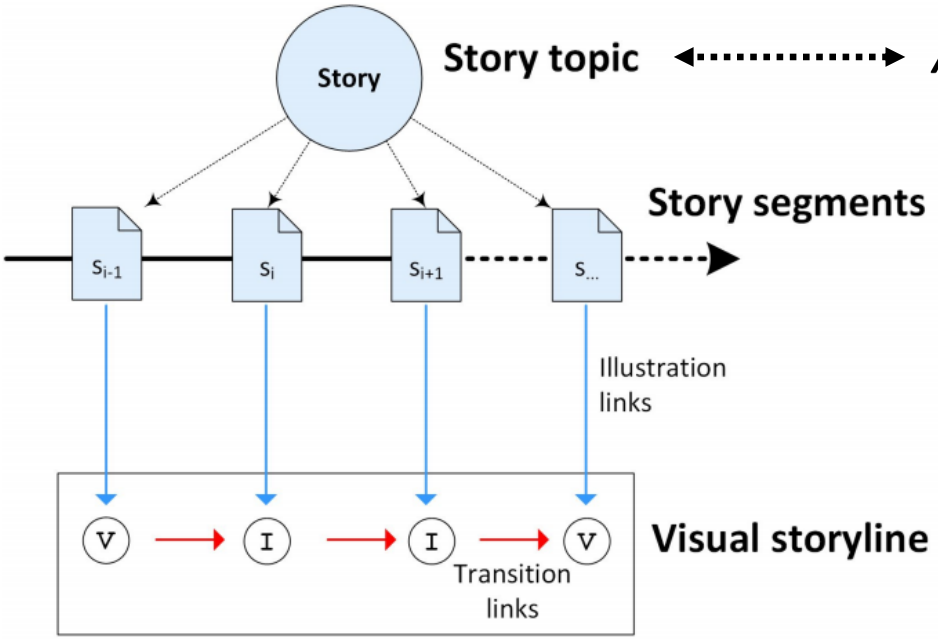
- **Visual storylines** are consistently used in news media to present information to the reader.

- In the newsroom, it is the job of the news editor to **find relevant images/videos** that illustrate specific stories and **organize them** in a **semantically**, **visually coherent** and **appealing fashion**, to create visual storylines.



Context and motivation

- The goal of the Social-media video story-telling linking is to **automatically illustrate a news story with social-media visual content**



***s*$_1$:** *Cyclist crash*

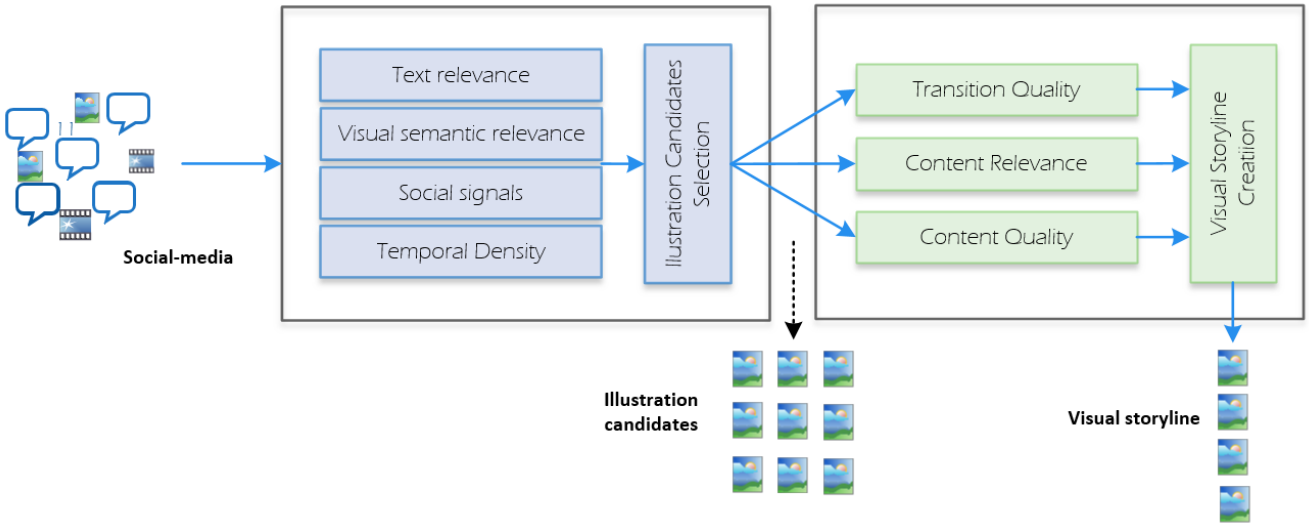
***s*$_2$:** *Bad weather*

***s*$_3$:** *Mechanical Problems*

***s*$_4$:** *Geraint Thomas forced to abandon race after crash*

Approach

- We propose a storyline illustration framework, leveraging on two components:
  - A component tasked with **retrieving relevant content**.
  - A component tasked with the **organization** of the retrieved relevant content **into visually coherent sequence**.



1 - Retrieving relevant content

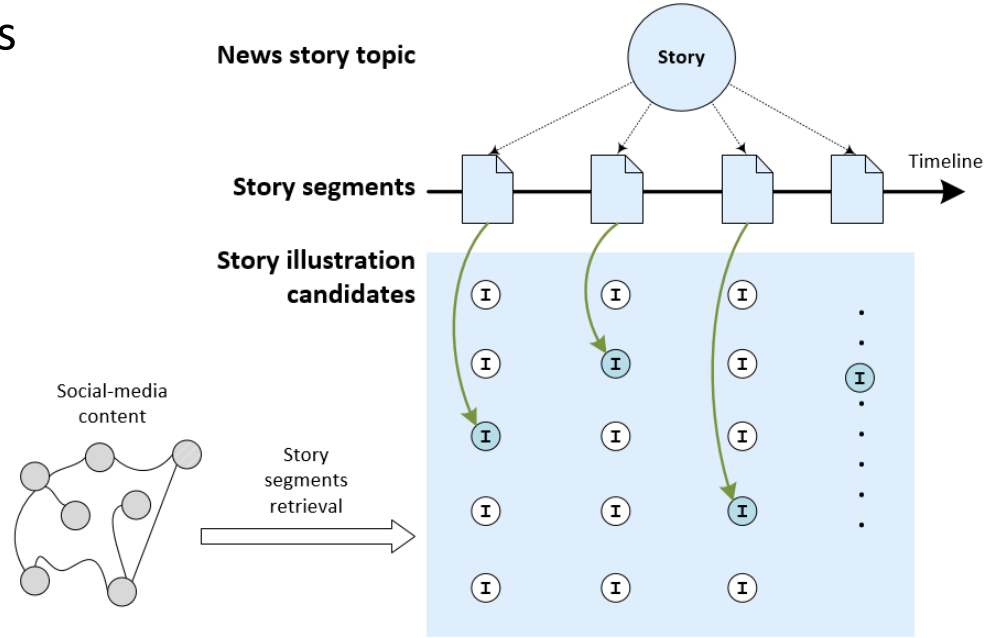
Combine the results of 5 retrieval models

Fuses them through **Reciprocal Rank Fusion**: weights each document with the inverse of its position on the rank.

$$RRFscore(d) = \sum_i \frac{1}{k + r_i(d)},$$

where k = 60

Exploit different retrieval models by favouring documents at the “top” of the rank.



Ranking relevant content

- Text retrieval (**TR**) using BM25 retrieval model.

Ranking relevant content

- Text retrieval (**TR**) using BM25 only.
- #Retweets (**RT**): TR and maximizing number of retweets.
- #Duplicated images (**Dup**): TR and maximizing number of duplicates.

Ranking relevant content

- Text retrieval (**TR**) using BM25 only.
- #Retweets (**RT**): TR and maximizing number of retweets.
- #Duplicated images (**Dup**): TR and maximizing number of duplicates.
- Concept Pool (**CP**): TR and extracting visual concepts, using a pre-trained VGG network, from the top-10 ranked tweets. Images are then re-ranked according to the number of visual concepts in the pool.
- Concept Query (**CQ**): TR and extracting visual concepts, from top-10 ranked tweets, creating a new query with those concepts. We fuse the two ranks using a rank fusion method (RRF), and the top ranked image is chosen.

Ranking relevant content

- Text retrieval (**TR**) using BM25 only.
- #Retweets (**RT**): TR and maximizing number of retweets.
- #Duplicated images (**Dup**): TR and maximizing number of duplicates.
- Concept Pool (CP): TR and extracting visual concepts, using a pre-trained VGG network, from the top-10 ranked tweets. Images are then re-ranked according to the number of visual concepts in the pool.
- Concept Query (CQ): TR and extracting visual concepts, from top-10 ranked tweets, creating a new query with those concepts. We fuse the two ranks using a rank fusion method (RRF), and the top ranked image is chosen.
- Temporal Modeling (**TM**): TR and creating a Kernel Density Estimation with the probability of a tweet being posted at a given date. The tweet that maximizes that probability is chosen.

2 - Illustrating storylines

A visual storyline is an ordered sequence of visual elements

Our rationale:

- From a non-computational perspective, transitions are characterized based on the relations between semantic and visual characteristics of adjacent images;
- We emulate this approach proposing a novel formalization of transition based on the **concept of distance**.

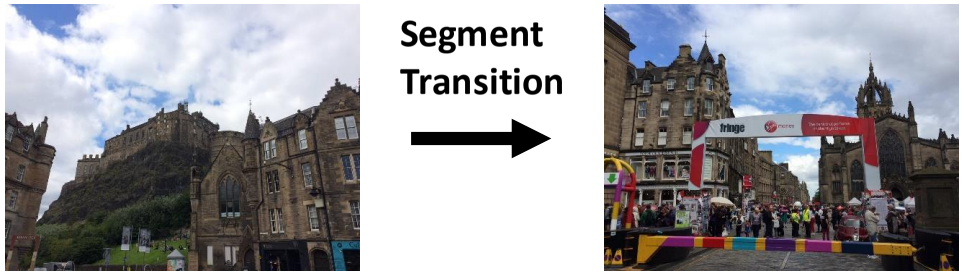
Given two sequential images a and b:

$$(\forall c \in C, distance_c(feature_c(a), feature_c(b)))$$

The chosen feature spaces should capture the semantic and visual characteristics

Inferring transition quality

A *Gradient Boosted Tree* regressor was trained to ***predict a rating given the transition distance of a pair***.



Input: Vector of pairwise distances, over different feature spaces, between each adjacent pair of images

Regressor model

Output: Predicted transition quality

Development data (2016 editions of EdFest and TDF) used for training: Annotated transitions (0 – *bad*, 1 – *acceptable*, 2 – *good*

Transition features considered

Input of the regressor model: Concatenation of pairwise distances, over 16 different visual feature spaces

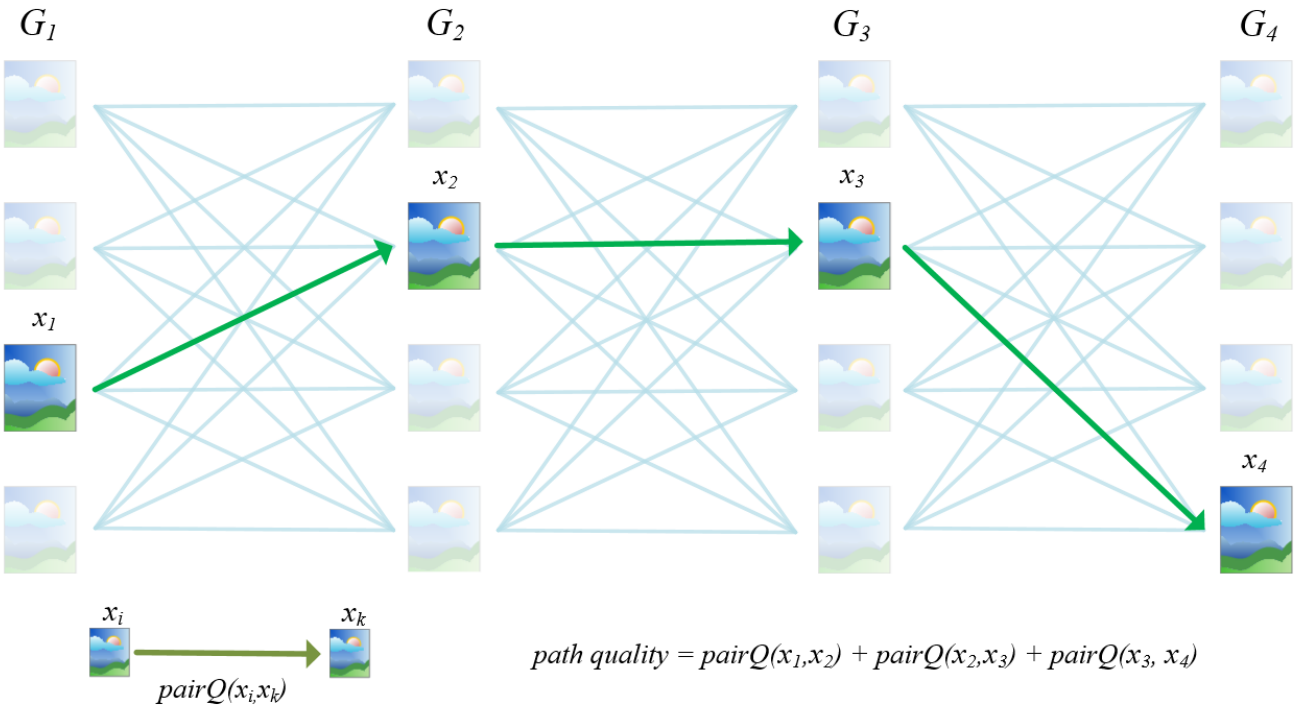
Quality Difference	$abs(f(S1) - f(S2))$	A positive real value representing the aesthetic quality of the image.
Quality Sum	$-abs(f(S1) + f(S2))$	A positive real value representing the aesthetic quality of the image.
Environment	$f(S1) = f(S2)$	If the image represents a place outdoors or indoors .
Faces	$abs(f(S1) - f(S2))$	The number of faces in the image.
Scene Attributes	$\#(f(S1) \cap f(S2))$	The characteristics of a scene described in individual words.
Scene Category	$\#(f(S1) \cap f(S2))$	The most probable locations of a scene described in individual words.
Color Correlogram	$jsd(f(S1) - f(S2))$	A 16 bins 3D color correlogram in LAB color space.
Heat Map	$\sum abs(f(S1) - f(S2))$	A heat map of informative parts of the image.
Color Histogram	$\sum abs(f(S1) - f(S2))$	A 16 bins 3D histogram in LAB color space.
CNN Dense	$\sum abs(f(S1) - f(S2))$	A thing extracted from the last layer of a neural network.
Color Moment	$euclidean(f(S1), f(S2))$	Color moment in LAB color space.
Entropy	$abs(f(S1) - f(S2))$	A positve real value representing the entropy.
Concepts	$\#(f(S1) \cap f(S2))$	A set of image concepts extracted using VGG16.
#Edges	$\sum abs(f(S1) - f(S2))$	A vector of three positions with the number of horizontal, vertical and diagonal edges, respectively.
Luminance	$abs(f(S1) - f(S2))$	A real value representing the luminance.
pHash	$\sum abs(f(S1) - f(S2))$	A Phash vector.

2 - Illustrating storylines

We propose four graph-based methods for storyline illustration:

Sequential without relevance (run 1): optimizes for the transition quality of adjacent elements pairs.

Sequential without relevance (run 1)



$$F_1 = \sum_{i=2}^{N} pairQ_1(i-1,i)$$

$$pairQ_1(i,k) = t_{i,k}$$

$t_{i,k}$ represents the normalized score of transition quality from image *i* to image *k*.

This score is attained through the use of a Gradient Boosted Trees regressor model

2 - Illustrating storylines

We propose four graph-based methods for storyline illustration:

Sequential without relevance (run 1): optimizes for the transition quality of adjacent elements pairs.

Sequential with relevance(run 2): leverages the transition quality of adjacent element pairs while taking into account relevance.

Sequential with relevance (run 2)

Directly optimise the task metric by approximating relevance and transitions quality:

$$Quality = 0.1 \cdot s_1 + \frac{0.9}{2(N-1)} \sum_{i=2}^{N} pairwiseQ(i)$$

$$pairwiseQ(i) = \underbrace{0.6 \cdot (s_i + s_{i-1})}_{\text{segments illustration}} + \underbrace{0.4 \cdot (s_{i-1} \cdot s_i + t_{i-1,i})}_{\text{transition}}$$

Here ***s*** represents the **normalized score of relevance** of an image to the segment it illustrates.

This score is attained through the use of the retrieval model described previously

2 - Illustrating storylines

We propose four graph-based methods for storyline illustration:

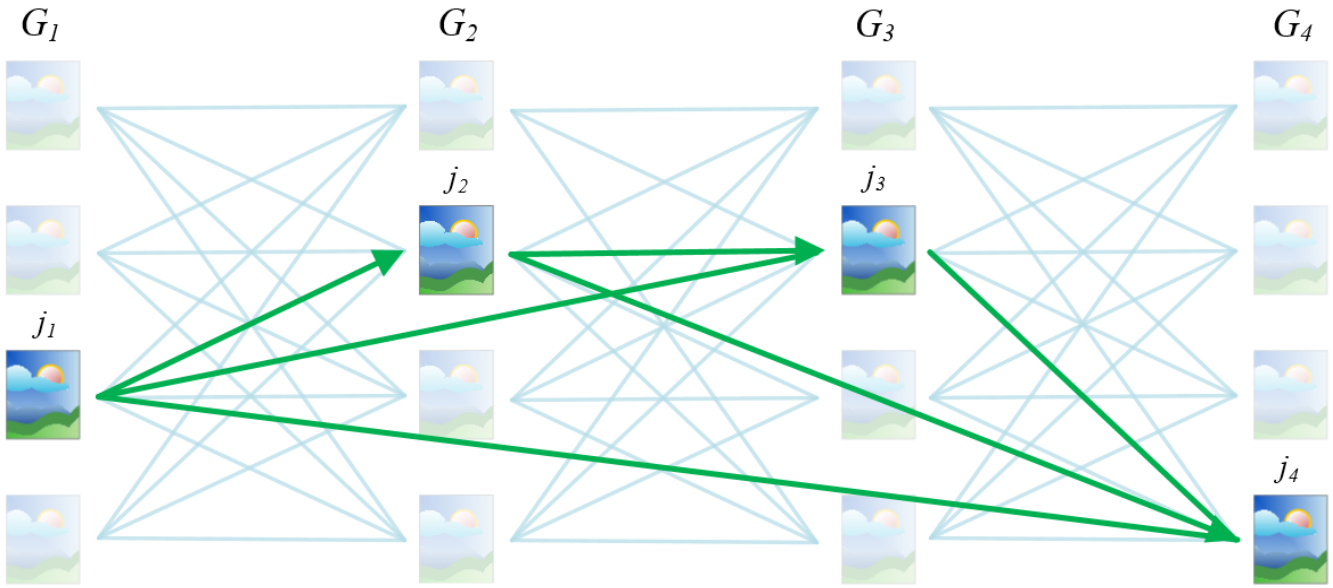
Sequential without relevance (run 1): optimizes for the transition quality of adjacent elements pairs.

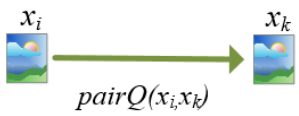
Sequential with relevance (run 2): leverages the transition quality of adjacent element pairs while taking into account relevance.

Fully connected without relevance (run 3): optimizes for transition quality between all pairs of images in the storyline.

Fully connected without relevance (run 3)

Optimise for transitions quality, for full sequences





$$path\ quality = pairQ(x_1, x_2) + pairQ(x_1, x_3) + pairQ(x_1, x_4) + pairQ(x_2, x_3) + pairQ(x_2, x_4) + pairQ(x_3, x_4)$$

$$F_3 = \sum_{i=1}^{N} \sum_{k \in \{2<=k<=N \wedge k \neq i\}} pairQ_1(i,k)$$

$$pairQ_1(i,k) = t_{i,k}$$

2 - Illustrating storylines

We propose four graph-based methods for storyline illustration:

Sequential without relevance (run 1): optimizes for the transition quality of adjacent elements pairs.
Sequential with relevance (run 2): leverages the transition quality of adjacent element pairs while taking into account relevance.
Fully connected without relevance (run 3): optimizes for transition quality between all pairs of images in the storyline.
Fully connected with relevance (run 4): leverages transition quality between all pairs of images in the storyline as well as relevance.

Fully connected with relevance (run 4)

Again, directly optimise the task metric by approximating relevance and transitions quality:

$$Quality = 0.1 \cdot s_1 + \frac{(0.9)}{2(N-1)} \sum_{i=2}^{N} segmentQ(i)$$

$$segmentQ(i) = \underbrace{0.6 \cdot (s_i)}_{\text{segments illustration}} + \underbrace{0.4 \cdot \sum_{k \in \{1<=k<=N \wedge k \neq i\}} (s_k \cdot s_i + t_{i,k})}_{\text{transitions}}$$

Results - Illustration Quality

Illustration quality metric:

$$pairwiseQ(i) = \underbrace{\beta \cdot (s_i + s_{i-1})}_{\text{segments illustration}} + \underbrace{(1-\beta) \cdot (s_{i-1} \cdot s_i + t_{i-1})}_{\text{transition}}$$

$$Quality = \alpha \cdot s_1 + \frac{(1-\alpha)}{2(N-1)} \sum_{i=2}^{N} pairwiseQ(i)$$

Edinburgh Festival 2017 Topics	run 1	ns_sequential_without_relevance	0.376333
	run 2	ns_sequential_with_relevance	0.360444
	run 3	**ns_fully_connected_without_relevance**	**0.402111**
	run 4	ns_fully_connected_with_relevance	0.300556
Tour de France 2017 Topics	run 1	ns_sequential_without_relevance	0.483667
	run 2	ns_sequential_with_relevance	0.462889
	run 3	**ns_fully_connected_without_relevance**	**0.554167**
	run 4	ns_fully_connected_with_relevance	0.506111

Results – Qualitative Analysis

(Run 4) Fully Connected with Relevance – Street Performances

The Edinburgh Festival is home to one of the most unique celebrations of arts

✔ Relevant

Street circus is a popular attraction at Edinburgh Festival with several artists such as unicycle jugglers

✔ Relevant

Street circus is full of colorful artists

✔ Relevant

Bagpipes

✔ Relevant

Results – Qualitative Analysis

(Run 3) Fully Connected without Relevance – Street Performances

The Edinburgh Festival is home to one of the most unique celebrations of arts

✓ Relevant

Street circus is a popular attraction at Edinburgh Festival with several artists such as unicycle jugglers

✓ Relevant

Street circus is full of colorful artists

✓ Relevant

Bagpipes

✓ Relevant

Results – Qualitative Analysis

(Run 3) Fully Connected without Relevance - Gastronomy at Edinburgh Festival

Pizzas

✔ Relevant

Hamburgers

✗ Not Relevant

Desserts

✔ Relevant

Drinks

✔ Relevant

Results – Qualitative Analysis

(Run 3) Fully Connected without Relevance – EdFest can be tiring

Crowds on the streets

✔ Relevant

People queueing

✔ Relevant

People standing watching a show

✔ Relevant

People laying down

X Not Relevant

Results – Qualitative Analysis

(Run 3) Fully Connected without Relevance – Scottish Elements

Bagpipes

✔ Relevant

Food and drinks

X Not Relevant

Traditional outfits

X Not Relevant

Military parade

X Not Relevant

Conclusions

We proposed a framework to **computationally emulate transition quality assessment**, by leveraging on a large set of feature spaces, each capturing different aspects

The proposed **regressor model contributed to the transitions quality** of story illustrations

With respect to **retrieval of relevant content**:

- Retrieval component needs to be improved. Consider using a cross-modal space, obtained by training on external data.
- Our relevance estimation model (run 2 and 4), based on retrieval models' scores, needs to be improved.%

Thank you